

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2001, 03:04:44 ; Search time 125.73 seconds

(without alignments)  
6518.961 Million cell updates/sec

Title: US-09-403-269-12\_COPY\_1\_1404  
Perfect score: 1404  
Sequence: 1 tccaagctgattctcatag.....gcagccggcagaacacac 1404

Scoring table:

IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

N\_Geneseq\_0401.\*  
1: /SID2/gcgdata/geneseq/geneseqn/NA1980.DAT.\*  
2: /SID2/gcgdata/geneseq/geneseqn/NA1981.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseqn/NA1982.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseqn/NA1983.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseqn/NA1984.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseqn/NA1985.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseqn/NA1986.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseqn/NA1987.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseqn/NA1988.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseqn/NA1989.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseqn/NA1990.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseqn/NA1991.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseqn/NA1992.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseqn/NA1993.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseqn/NA1994.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseqn/NA1995.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseqn/NA1996.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseqn/NA1997.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseqn/NA1998.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseqn/NA1999.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseqn/NA2000.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseqn/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1404	100.0	3085	19	V62688
2	35.6	2.5	237326	19	V57903
3	35	2.5	1366	20	V27743
4	35	2.5	5059	20	X84332
5	34.4	2.5	2248	21	C69113
6	33.8	2.4	571	20	X20393
7	33.8	2.4	9542	20	X20260
8	33.6	2.4	283	21	C94048
9	33.6	2.4	1231	21	C39833
10	33.6	2.4	1392	19	V44080
11	33.6	2.4	1534	16	O92554

C	12	33	2.4	1644	20	X60775
C	13	32.6	2.3	3970	21	A97911
C	14	32.6	2.3	6204	21	A97917
C	15	32.6	2.3	7387	21	A97914
C	16	32.6	2.3	9321	21	A97904
C	17	32.6	2.3	4999	20	E23903
C	18	32.4	2.3	72604	20	T10752
C	19	32	2.3	1334	21	A39409
C	20	32	2.3	1958	19	V52493
C	21	32	2.3	2423	21	C98795
C	22	31.8	2.3	1661	17	T45120
C	23	31.8	2.3	1561	19	V03875
C	24	31.6	2.3	1339	19	V44234
C	25	31.6	2.3	624	21	F07873
C	26	31.6	2.3	2547	21	A74993
C	27	31.6	2.3	2773	21	E65009
C	28	31.6	2.3	4473	22	F44671
C	29	31.4	2.2	1238	21	E52513
C	30	31.4	2.2	1291	21	C39019
C	31	31.4	2.2	1299	21	A49460
C	32	31.4	2.2	1330	21	C36038
C	33	31.4	2.2	1333	21	C49301
C	34	31.4	2.2	1365	20	X85015
C	35	31.4	2.2	1447	20	X13590
C	36	31.4	2.2	1486	21	F18228
C	37	31.4	2.2	1627	16	O95058
C	38	31.4	2.2	1627	16	V37721
C	39	31.4	2.2	1666	16	O95059
C	40	31.4	2.2	1666	19	V37722
C	41	31.4	2.2	3463	21	C73782
C	42	31.4	2.2	51259	18	X83007
C	43	31.4	2.2	1664976	19	V21209
C	44	31.2	2.2	1551	21	C77099
C	45	31.2	2.2	1646	9	N82246

## ALIGNMENTS

RESULT	ID	CDNA	BP
1	V62688	standard	3085
1	V62688	standard	3085
15-FEB-1999	(first entry)		
Bovine glucuronyl C5-epimerase cDNA.			
Glucuronyl C5-epimerase; cattle; D-glucuronic acid; L-iduronic acid; heparin; heparan sulphate; ss.			
Bos taurus.			
Key	Location/Qualifiers		
CDS	73..1407		
WT	1998-583655/49.		
P-PSDB	W79263.		

Soybean cytochrome  
L. mesenteroides a  
L. mesenteroides a  
L. mesenteroides a  
L. mesenteroides a  
Human LOBO homolog  
Genomic sequence c  
Rice SYR2 homolog  
Streptococcus pneu  
Human pancreatic c  
Htm4 gene. Homo s  
Lettuce resistance  
Fusarium venenatum  
CDNA encoding a hu  
Membrane-bound pro  
Novel protein kina  
Human secreted pro  
Arabidopsis thalia  
Human CD2 cytoplas  
Arabidopsis thalia  
Arabidopsis thalia  
Human secreted pro  
Enterococcus faeca  
Lung cancer associ  
Chicken TEF-1A cDN  
Recombinant transc  
Human TEF-1B cDN  
Recombinant transc  
Human OREF ORF2937  
Partial mouse WRN  
Methanococcus jann  
Human OREF ORF2654  
Rice storage prote

XX DNA sequence coding for mammalian glucuronyl C5-epimerase and  
 PT functional derivatives - capable of converting D-glucuronic acid to  
 PT L-iduronic acid in the synthesis of heparin and heparan sulphate  
 XX  
 PS Claim 2; Page 18-19; 26pp; English.

CC This cDNA sequence includes a claimed coding region that encodes  
 CC bovine glucuronyl C5-epimerase (see W9263), an enzyme that  
 CC catalyses the conversion of D-glucuronic acid (GlcA) to L-iduronic  
 CC acid (IdoA). To isolate the sequence, highly purified epimerase  
 CC from bovine liver was subjected to digestion with a lysine-specific  
 CC protease. A DNA probe was generated from bovine liver cDNA by PCR  
 CC using primers (see W62689-91) based on an isolated peptide (see  
 CC W9270). This was used to screen a bovine lung lambda gt10 library,  
 CC and a hybridising clone was identified and sequenced. An additional  
 CC 12 bp of 5' sequence was obtained from a clone isolated from the  
 CC library by PCR using an epimerase-specific primer (see W62692).  
 CC The invention relates to isolated or recombinant DNA sequences for  
 CC a mammalian (including human) glucuronyl C5-epimerase or its  
 CC functional derivative. Recombinant expression vectors and  
 CC transformed host cells are also claimed. The nucleic acid and  
 CC vector can be used for the recombinant production of the enzyme.  
 CC Glucuronyl C5-epimerase is useful for converting GlcA to IdoA in  
 CC the biosynthesis of heparin and heparan sulphate.

Sequence 3085 BP: 903 A; 625 C; 654 G; 903 T; 0 other;

Query Match 100.0%; Score 1404; DB 19; Length 3085;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tccaagctgaattcctcagctatcccaagctctatgacagagagccctatcacct 60  
 DB 1 tccaagctgaattcctcagctatcccaagctctatgacagagagccctatcacct 60  
 QY 61 gatgctgtcttctgctcttgaagctacaatgtggaagctccgagacagatcaatgctc 120  
 DB 61 gatgctgtcttctgctcttgaagctacaatgtggaagctccgagacagatcaatgctc 120  
 QY 61 gatgctgtcttctgctcttgaagctacaatgtggaagctccgagacagatcaatgctc 120  
 DB 61 gatgctgtcttctgctcttgaagctacaatgtggaagctccgagacagatcaatgctc 120  
 QY 121 ataatgctgtgctggaagcttacccttaccagctgagaccccaagctattctac 180  
 DB 121 ataatgctgtgctggaagcttacccttaccagctgagaccccaagctattctac 180  
 QY 121 ataatgctgtgctggaagcttacccttaccagctgagaccccaagctattctac 180  
 DB 121 ataatgctgtgctggaagcttacccttaccagctgagaccccaagctattctac 180  
 QY 181 ccaatcagatctcagacagatggttaagctcaactacagcaagaatctaatgtaaaacc 240  
 DB 181 ccaatcagatctcagacagatggttaagctcaactacagcaagaatctaatgtaaaacc 240  
 QY 181 ccaatcagatctcagacagatggttaagctcaactacagcaagaatctaatgtaaaacc 240  
 DB 181 ccaatcagatctcagacagatggttaagctcaactacagcaagaatctaatgtaaaacc 240  
 QY 241 cctcatatagagctatatagaagaagcagagacagagacaaagacagccatgac 300  
 DB 241 cctcatatagagctatatagaagaagcagagacagagacaaagacagccatgac 300  
 QY 241 cctcatatagagctatatagaagaagcagagacagagacaaagacagccatgac 300  
 DB 241 cctcatatagagctatatagaagaagcagagacagagacaaagacagccatgac 300  
 QY 301 tggactctgcccagagctgcttactgctgagctgctgctgataagctcaagattaccaat 360  
 DB 301 tggactctgcccagagctgcttactgctgagctgctgctgataagctcaagattaccaat 360  
 QY 301 tggactctgcccagagctgcttactgctgagctgctgctgataagctcaagattaccaat 360  
 DB 301 tggactctgcccagagctgcttactgctgagctgctgctgataagctcaagattaccaat 360  
 QY 361 gttaaaagctcattgctccagaaacagtggaagtgatctcttggtaacttggggaacaa 420  
 DB 361 gttaaaagctcattgctccagaaacagtggaagtgatctcttggtaacttggggaacaa 420  
 QY 361 gttaaaagctcattgctccagaaacagtggaagtgatctcttggtaacttggggaacaa 420  
 DB 361 gttaaaagctcattgctccagaaacagtggaagtgatctcttggtaacttggggaacaa 420  
 QY 421 aaaaatttatattatcttacttgcctcaagttcttaacaatggaagcgctgtctggtt 480  
 DB 421 aaaaatttatattatcttacttgcctcaagttcttaacaatggaagcgctgtctggtt 480  
 QY 421 aaaaatttatattatcttacttgcctcaagttcttaacaatggaagcgctgtctggtt 480  
 DB 421 aaaaatttatattatcttacttgcctcaagttcttaacaatggaagcgctgtctggtt 480  
 QY 481 ctgagagacagaaagaatcagctctctcactgataatgcttcaaataccagctca 540  
 DB 481 ctgagagacagaaagaatcagctctctcactgataatgcttcaaataccagctca 540  
 QY 481 ctgagagacagaaagaatcagctctctcactgataatgcttcaaataccagctca 540  
 DB 481 ctgagagacagaaagaatcagctctctcactgataatgcttcaaataccagctca 540  
 QY 541 attgcttttaagaagaagacatactatgagatcggtggccagacatcatgagacaa 600  
 DB 541 attgcttttaagaagaagacatactatgagatcggtggccagacatcatgagacaa 600

QY 601 gttaccggagctggtcacttactcaggaagaagtggtgtcttccacacaaagct 660  
 DB 601 gttaccggagctggtcacttactcaggaagaagtggtgtcttccacacaaagct 660  
 QY 661 gtaagccacacagaataatgcccagaagtgctgtaagttgtagtcgaagggagggc 720  
 DB 661 gtaagccacacagaataatgcccagaagtgctgtaagttgtagtcgaagggagggc 720  
 QY 661 gtaagccacacagaataatgcccagaagtgctgtaagttgtagtcgaagggagggc 720  
 DB 661 gtaagccacacagaataatgcccagaagtgctgtaagttgtagtcgaagggagggc 720  
 QY 721 ttccttgacacattacatctctaccacagccacatgctgctcttcttgcctcagct 780  
 DB 721 ttccttgacacattacatctctaccacagccacatgctgctcttcttgcctcagct 780  
 QY 721 ttccttgacacattacatctctaccacagccacatgctgctcttcttgcctcagct 780  
 DB 721 ttccttgacacattacatctctaccacagccacatgctgctcttcttgcctcagct 780  
 QY 781 gactggtgttggaagcagcagatggaagagcgctggtgcgattatggtgacccgtaag 840  
 DB 781 gactggtgttggaagcagcagatggaagagcgctggtgcgattatggtgacccgtaag 840  
 QY 781 gactggtgttggaagcagcagatggaagagcgctggtgcgattatggtgacccgtaag 840  
 DB 781 gactggtgttggaagcagcagatggaagagcgctggtgcgattatggtgacccgtaag 840  
 QY 841 ttaagggaagagctcagctctttagagcagagtggtgactccgacatggccaaaggcaa 900  
 DB 841 ttaagggaagagctcagctctttagagcagagtggtgactccgacatggccaaaggcaa 900  
 QY 841 ttaagggaagagctcagctctttagagcagagtggtgactccgacatggccaaaggcaa 900  
 DB 841 ttaagggaagagctcagctctttagagcagagtggtgactccgacatggccaaaggcaa 900  
 QY 901 gccatttctaatatgacagagcgtctctcttcttcaaaaagacatatattctcaatca 960  
 DB 901 gccatttctaatatgacagagcgtctctcttcttcaaaaagacatatattctcaatca 960  
 QY 901 gccatttctaatatgacagagcgtctctcttcttcaaaaagacatatattctcaatca 960  
 DB 901 gccatttctaatatgacagagcgtctctcttcttcaaaaagacatatattctcaatca 960  
 QY 961 gctttaaggcaaacagcccttacaagttctgctcagagcagcatgagtgagctgtg 1020  
 DB 961 gctttaaggcaaacagcccttacaagttctgctcagagcagcatgagtgagctgtg 1020  
 QY 961 gctttaaggcaaacagcccttacaagttctgctcagagcagcatgagtgagctgtg 1020  
 DB 961 gctttaaggcaaacagcccttacaagttctgctcagagcagcatgagtgagctgtg 1020  
 QY 1021 ttatgataaacatgactggtgatggaagatctacactacactagctcttcttctta 1080  
 DB 1021 ttatgataaacatgactggtgatggaagatctacactacactagctcttcttctta 1080  
 QY 1021 ttatgataaacatgactggtgatggaagatctacactacactagctcttcttctta 1080  
 DB 1021 ttatgataaacatgactggtgatggaagatctacactacactagctcttcttctta 1080  
 QY 1081 aatgcttattgattcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1140  
 DB 1081 aatgcttattgattcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1140  
 QY 1081 aatgcttattgattcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1140  
 DB 1081 aatgcttattgattcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1140  
 QY 1141 ctgaggaagaagcagagctctcttcttcttcttcttcttcttcttcttcttcttctt 1200  
 DB 1141 ctgaggaagaagcagagctctcttcttcttcttcttcttcttcttcttcttcttctt 1200  
 QY 1141 ctgaggaagaagcagagctctcttcttcttcttcttcttcttcttcttcttcttctt 1200  
 DB 1141 ctgaggaagaagcagagctctcttcttcttcttcttcttcttcttcttcttcttctt 1200  
 QY 1201 cctgtttagagacagctgctcaggaacacatcactagaccccgacatcagctgtgcat 1260  
 DB 1201 cctgtttagagacagctgctcaggaacacatcactagaccccgacatcagctgtgcat 1260  
 QY 1201 cctgtttagagacagctgctcaggaacacatcactagaccccgacatcagctgtgcat 1260  
 DB 1201 cctgtttagagacagctgctcaggaacacatcactagaccccgacatcagctgtgcat 1260  
 QY 1261 gccccaacctggtccgtctggaactatcacacacacacacacacacacacacacacac 1320  
 DB 1261 gccccaacctggtccgtctggaactatcacacacacacacacacacacacacacacac 1320  
 QY 1261 gccccaacctggtccgtctggaactatcacacacacacacacacacacacacacacac 1320  
 DB 1261 gccccaacctggtccgtctggaactatcacacacacacacacacacacacacacacac 1320  
 QY 1321 agcacatgagagctcccaatcttcaagaattgtcagaagtggaagagctacact 1380  
 DB 1321 agcacatgagagctcccaatcttcaagaattgtcagaagtggaagagctacact 1380  
 QY 1321 agcacatgagagctcccaatcttcaagaattgtcagaagtggaagagctacact 1380  
 DB 1321 agcacatgagagctcccaatcttcaagaattgtcagaagtggaagagctacact 1380  
 QY 1381 aaaggcagccgagcaagacacac 1404  
 DB 1381 aaaggcagccgagcaagacacac 1404

# RESULT 2

V57903/C  
 ID V57903 standard; DNA: 237326 BP.

XX V57903;

XX 21-DEC-1998 (first entry)

XX Hereditary haemochromatosis subregion from an HH affected individual.

XX Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;

XX diagnosis; Iron metabolism; NPT3; NPT4; Roret; BTF1; BTF2; BTF3;

XX BTF4; BTF5; milk protein; Lupus; Sjogren's syndrome; hypophosphatemia;

XX type 1 sodium transport gene; SS.

XX Homo sapiens.



RESULT	ID	Location/Qualifiers	base numbering
4	X84332		
XX	X84332 standard; DNA; 5059 BP.		
AC	X84332;		
XX	08-SEP-1999 (first entry)		
DT			
XX	Stealth virus nucleic acid clone, SEQ ID NO: 24.		
DE	Stealth virus; detection; diagnosis; infection; ss.		
XX			
KM	Stealth virus.		
XX			
OS			
Key			
PH	misc_difference 3605		
FT	/*tag-	a	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note-		
FT			
FT	misc_difference 3610		
FT	/*tag-	b	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note-		
FT			
FT	misc_difference 3615		
FT	/*tag-	c	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note-		
FT			
FT	misc_difference 3630		
FT	/*tag-	d	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note-		
FT			
FT	misc_difference 3631		
FT	/*tag-	e	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note-		
FT			
FT	misc_difference 3638		
FT	/*tag-	f	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note-		
FT			
FT	misc_difference 3641		
FT	/*tag-	g	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note-		
FT			
FT	misc_difference 3642		
FT	/*tag-	h	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note-		
FT			
FT	misc_difference 3657		
FT	/*tag-	i	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note-		
FT			
FT	misc_difference 3659		
FT	/*tag-	j	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note-		
FT			
FT	misc_difference 3681		
FT	/*tag-	k	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note-		
FT			
FT	misc_difference 3698		
FT	/*tag-	l	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note-		
FT			

[illegible]

```

FT      /*tag- aa
FT      /note- "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc-difference 3812
FT      /*tag- ab
FT      /note- "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc-difference 3813
FT      /*tag- ac
FT      /note- "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc-difference 3818
FT      /*tag- ad
FT      /note- "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc-difference 3819
FT      /*tag- ae
FT      /note- "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc-difference 3824
FT      /*tag- af
FT      /note- "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc-difference 3827
FT      /*tag- ag
FT      /note- "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc-difference 3841
FT      /*tag- ah
FT      /note- "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc-difference 3853
FT      /*tag- ai
FT      /note- "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc-difference 3857
FT      /*tag- aj
FT      /note- "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc-difference 3860
FT      /*tag- ak
FT      /note- "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc-difference 3863
FT      /*tag- al
FT      /note- "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
Query Match 2.5%; Score 35; DB 20; Length 5059;
Best Local Similarity 15.7%; Pred. No. 3;
Matches 53; Conservative 120; Mismatches 164; Indels 0; Gaps 0;

```

```

OY 346 tcaagattccacaaagtlttaacagcttcctccagaaacagtgagtgatctc 405
DB 2479 tthcmvgnmamsarsbwnscsnltgmthynccadstranhcmvnrscntdsbas 2538
OY 406 caactgggagacaaagaattattattcattgcactgaagcttaacaatga 465
DB 2539 arandubntaycdsrstasrtrnsrshncasatshcmvncanmadtnartncdngs 2598
OY 466 agcgtctcgtgtctcttgagacgacagaagaatc 502
DB 2599 ncsntrsnthbratryadadtdadstranmtlmsnc 2635

RESULT 5
C69113
ID C69113 standard; DNA: 2248 BP.
AC C69113;
DT 31-JAN-2001 (first entry)
XX
DE Human secreted protein gene 9 clone HTEOF33.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ss.
XX
OS Homo sapiens.
XX
PN WO200055371-A1.
XX
PD 21-SEP-2000.
XX
PF 16-MAR-2000; 2000WO-US06783.
XX
PR 18-MAR-1999; 99US-0125055.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Ni J, Ebner R, Rosen CA, Shi Y, Birse C, Florence K;
PI Komatsoulis G, Lafleur DW, Moore PA, Olsen HS, Young PE;
XX
DR MPI: 2000-594448/56.
XX
P-PSDB: B38013.
XX
PT New nucleic acid molecules encoding 27 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
PS Claim 1: Page 387; 453pp; English.
XX
CC The invention relate to the isolation of genes C69084-C69119 encoding
CC 27 human secreted proteins B37984-B38019. The genes can be used to
CC generate fusion proteins by linking to the gene for the human
CC immunoglobulin G Fc portion (C69075) for increasing the stability of
CC the fusion protein as compared to the human protein only. The genes and
CC proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated
CC from a range of human tissues disclosed in the specification. The
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC and (f) infectious diseases such as viral, bacterial, fungal and
CC parasitic infections.
XX

```







PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136752.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139482.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142920.  
PR 13-JUL-1999; 99US-0142977.  
PR 14-JUL-1999; 99US-0143342.  
PR 15-JUL-1999; 99US-0143624.  
PR 16-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144336.  
PR 20-JUL-1999; 99US-0144337.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155119.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157866.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160960.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.



conservative 0; Mismatches 54; Indels 0; Gaps 0

Matches	66; Conservative	0; Mismatches
Matches	66; Conservative	0; Mismatches



XX AC A97917;  
 XX 19-JAN-2001 (first entry)  
 XX L. mesenteroides alternan sucrose insertion sequence DNA.  
 XX DE  
 XX KW Alternan sucrose; glucosyltransferase; fructose; cosmetic; foodstuff;  
 XX KW syrup; ds.  
 XX OS  
 XX PN Leuconostoc mesenteroides.  
 XX DE19905069-A1.  
 XX PD 10-AUG-2000.  
 XX PF 08-FEB-1999; 99DE-1005069.  
 XX PR 08-FEB-1999; 99DE-1005069.  
 XX PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.  
 XX PI (PLAC) MAX PLANCK GES FORDERUNG WISSENSCHAFTEN.  
 XX PI Kossmann J, Welsh T, Quanz M, Knuth K;  
 XX WPI: 2000-550294/51.  
 XX DR  
 XX PT New nucleic acid encoding recombinant leuconostoc mesenteroides  
 XX PT alternan sucrose protein and methods of alternan and fructose  
 XX PS production  
 XX PS Example 4; Page 51-54; 64pp; German.  
 CC This invention describes a novel nucleic acid molecule (1) encoding an  
 CC alternan sucrose (E.C. 2.4.1.140 - an enzyme, that belongs to the  
 CC glucosyltransferase group) The recombinant, purified alternan sucrose  
 CC gene is useful for the fermentative production of alternan (a  
 CC carbohydrate) and/or fructose by secreting the enzyme into a  
 CC saccharose-containing culture medium. Alternatively, the enzyme is  
 CC contacted with a saccharose-containing solution. The enzyme is  
 CC fructose is then isolated from the medium. Cosmetic products or  
 CC foodstuffs containing alternan can be produced. Recombinant production of  
 CC alternan sucrose is advantageous as it provides a cost effective means of  
 CC producing fructose for high fructose containing syrups, production of  
 CC starch. This sequence represents a leuconostoc mesenteroides alternan  
 CC sucrose protein encoding DNA insertion sequence which is described in the  
 CC method of the invention.  
 CC SQ Sequence 6204 BP: 2101 A; 1038 C; 1354 G; 1711 T; 0 other;

Query Match 2.3%; Score 32.6; DB 21; Length 6204;  
 Best Local Similarity 52.6%; Pred. No. 20;  
 Matches 71; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

OY 426 ttattattattacattgacctgaagttcttaacaatgagagcgtctggtgttcgga 485  
 DB 2787 tttgtgttaataaccctgaccttaagttaacaatgataagaccattacattgatatgg 2846  
 OY 486 gacgacagaagaatcagctcttctacgtatgataatgataatcagcagctatgac 545  
 DB 2847 aaaggcgacataagatacattacgcgtcttagtatatacaatgatacagcagatgga 2906  
 OY 546 tttaagaagaaga 560  
 DB 2907 ttttatgatagtga 2921

RESULT 15  
 ID A97914  
 XX A97914 standard; DNA: 7387 BP.

AC A97914;  
 XX 19-JAN-2001 (first entry)  
 XX L. mesenteroides alternan sucrose 3'UTR.  
 XX DE  
 XX KW Alternan sucrose; glucosyltransferase; fructose; cosmetic; foodstuff;  
 XX KW syrup; ds.  
 XX OS  
 XX PN Leuconostoc mesenteroides.  
 XX DE19905069-A1.  
 XX PD 10-AUG-2000.  
 XX PF 08-FEB-1999; 99DE-1005069.  
 XX PR 08-FEB-1999; 99DE-1005069.  
 XX PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.  
 XX PI (PLAC) MAX PLANCK GES FORDERUNG WISSENSCHAFTEN.  
 XX PI Kossmann J, Welsh T, Quanz M, Knuth K;  
 XX WPI: 2000-550294/51.  
 XX DR  
 XX PT New nucleic acid encoding recombinant leuconostoc mesenteroides  
 XX PT alternan sucrose protein and methods of alternan and fructose  
 XX PS production  
 XX PS Example 2; Page 46-50; 64pp; German.  
 CC This invention describes a novel nucleic acid molecule (1) encoding an  
 CC alternan sucrose (E.C. 2.4.1.140 - an enzyme, that belongs to the  
 CC glucosyltransferase group) The recombinant, purified alternan sucrose  
 CC gene is useful for the fermentative production of alternan (a  
 CC carbohydrate) and/or fructose by secreting the enzyme into a  
 CC saccharose-containing culture medium. Alternatively, the enzyme is  
 CC contacted with a saccharose-containing solution. The enzyme is  
 CC fructose is then isolated from the medium. Cosmetic products or  
 CC foodstuffs containing alternan can be produced. Recombinant production of  
 CC alternan sucrose is advantageous as it provides a cost effective means of  
 CC producing fructose for high fructose containing syrups, production of  
 CC starch. This sequence represents a leuconostoc mesenteroides alternan  
 CC sucrose 3'-UTR DNA fragment which is described in the method of  
 CC the invention.  
 CC SQ Sequence 7387 BP: 2469 A; 1218 C; 1581 G; 2119 T; 0 other;

Query Match 2.3%; Score 32.6; DB 21; Length 7387;  
 Best Local Similarity 52.6%; Pred. No. 22;  
 Matches 71; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

OY 426 ttattattattacattgacctgaagttcttaacaatgagagcgtctggtgttcgga 485  
 DB 3458 tttgtgttaataaccctgaccttaagttaacaatgataagaccattacattgatatgg 3517  
 OY 486 gacgacagaagaatcagctcttctacgtatgataatgataatcagcagctatgac 545  
 DB 3518 aaaggcgacataagatacattacgcgtcttagtatatacaatgatacagcagatgga 3577  
 OY 546 tttaagaagaaga 560  
 DB 3578 ttttatgatagtga 3592

Fri May 25 10:28:25 2001

us-09-403-269-12\_copy\_1\_1404.rng

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2001, 02:22:24 ; Search time 67.79 Seconds

(without alignments)  
3616,200 Million cell updates/sec

Title: US-09-403-269-12\_COPY\_1\_1404  
Perfect score: 1404  
Sequence: 1 tccaagctgaatttcataag.....gcagccggcaagacacacac 1404

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

- 1: /cgn2-6/ptodata/2/ina/5A.COMB.seq:\*
- 2: /cgn2-6/ptodata/2/ina/5B.COMB.seq:\*
- 3: /cgn2-6/ptodata/2/ina/6A.COMB.seq:\*
- 4: /cgn2-6/ptodata/2/ina/6B.COMB.seq:\*
- 5: /cgn2-6/ptodata/2/ina/PCTUS.COMB.seq:\*
- 6: /cgn2-6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	37.4	2.7	7218	1 US-08-232-463-14	Sequence 14, Appl
C 2	35.4	2.5	474	4 US-09-276-531-29	Sequence 29, Appl
C 3	33	2.4	1644	4 US-08-948-564-5	Sequence 5, Appl
C 4	31.8	2.3	1661	1 US-08-318-492-3	Sequence 3, Appl
C 5	31.8	2.3	1661	1 US-08-707-340-3	Sequence 3, Appl
C 6	31.8	2.3	1661	1 US-08-994-578-3	Sequence 3, Appl
C 7	31.4	2.2	1627	1 US-08-615-170-2	Sequence 4, Appl
C 8	31.4	2.2	1666	1 US-08-615-170-4	Sequence 4, Appl
C 9	31.4	2.2	51259	2 US-08-781-891-209	Sequence 209, App
C 10	31	2.2	3639	2 US-08-737-524B-26	Sequence 26, Appl
C 11	30.8	2.2	4041	1 US-08-147-812-4	Sequence 4, Appl
C 12	30.8	2.2	4110	4 US-09-123-708-1	Sequence 1, Appl
C 13	30.8	2.2	4110	4 US-09-123-708-1	Sequence 1, Appl
C 14	30.8	2.2	4165	1 US-08-147-812-6	Sequence 1, Appl
C 15	30.6	2.2	1605	3 US-08-701-582D-1	Sequence 1, Appl
C 16	30.6	2.2	1710	3 US-09-253-912-1	Sequence 1, Appl
C 17	30.6	2.2	1710	4 US-09-082-039A-14	Sequence 14, Appl
C 18	30.6	2.2	176373	4 US-09-128-155-17	Sequence 17, Appl
C 19	30.4	2.2	1603	1 US-07-885-970A-19	Sequence 19, Appl
C 20	30.4	2.2	1603	1 US-08-298-687A-19	Sequence 19, Appl
C 21	30.4	2.2	1603	1 US-08-298-687A-19	Sequence 19, Appl
C 22	30.4	2.1	3164	1 US-08-188-228-49	Sequence 49, Appl
C 23	30	2.1	3164	1 US-08-332-643-43	Sequence 43, Appl
C 24	30	2.1	3164	1 US-08-332-643-43	Sequence 43, Appl
C 25	30	2.1	4380	1 US-07-582-945-1	Sequence 1, Appl
C 26	30	2.1	4380	2 US-08-453-141-1	Sequence 1, Appl
C 27	30	2.1	4380	4 US-08-293-314-1	Sequence 1, Appl

28	29.6	2.1	2924	3 US-09-357-073-1	Sequence 1, Appl
29	29.4	2.1	1473	2 US-08-602-725-31	Sequence 31, Appl
30	29.4	2.1	1630	3 US-08-468-856B-2	Sequence 2, Appl
31	29.4	2.1	1630	3 US-08-468-856B-2	Sequence 2, Appl
32	29.4	2.1	1886	1 US-07-891-942G-4	Sequence 4, Appl
33	29.4	2.1	2370	2 US-08-838-219B-19	Sequence 19, Appl
34	29.4	2.1	2370	4 US-09-233-336A-19	Sequence 19, Appl
35	29.4	2.1	2370	4 US-09-233-336A-19	Sequence 19, Appl
36	29.4	2.1	2911	2 US-08-588-983-11	Sequence 11, Appl
37	29.4	2.1	2911	2 US-08-588-983-11	Sequence 11, Appl
38	29.4	2.1	3461	3 US-08-468-856B-6	Sequence 6, Appl
39	29.4	2.1	3461	3 US-08-468-856B-6	Sequence 6, Appl
40	29.4	2.1	7920	4 US-09-080-983-2	Sequence 2, Appl
41	29.4	2.1	15500	4 US-09-080-983-1	Sequence 1, Appl
C 42	29.2	2.1	2778	1 US-08-202-054-1	Sequence 1, Appl
C 43	29.2	2.1	2778	1 US-08-446-923-1	Sequence 1, Appl
C 44	29	2.1	423	1 US-08-470-179-176	Sequence 176, App
C 45	29	2.1	1834	1 US-08-592-126-90	Sequence 90, Appl

## ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367

GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOXP2 VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE: 26-AUG-1991  
PRIORITY APPLICATION DATA:  
CLASSIFICATION: 435  
PRIOR APPLICATION NUMBER: US/07/935,313  
FILING DATE: 26-AUG-1991  
APPLICATION NUMBER: EP 91 114 300.6  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)836-4109  
TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-Fls  
US-08-232-463-14

Query Match 2.7%; Score 37.4; DB 1; Length 7218;  
 Best Local Similarity 6.2%; Pred. No. 0.13;  
 Matches 26; Conservative 207; Mismatches 188; Indels 0; Gaps 0;

7 ctgaattcctatgctatcccaagctatgacagagagccctatccatgctatgt 66  
 1485 CTGTAATTACTTCTATGCAACTAGTTAAGAGATAGAAGATTGTCACRRRRRRR 1426  
 67 gtgttattccttgaaggctacaatgtgaaagtcgagacagatcagtgatagt 126  
 1425 RRR 1366  
 127 ggggttaaggctacttattatcagacagtgaggcccaagctatcccaatc 186  
 1365 RRR 1306  
 187 cagattcagacagataggttaagctacacagacaagaatcctaactaataccctct 246.  
 1305 RRR 1246  
 247 atagaggtatataaacaagcagaagacagggacaagaacagcccaatgactgact 306  
 1245 RRR 1186  
 307 gtgccaagggctcttattatgctagtgctgataagctaacatgtttaa 366  
 1165 RRR 1126  
 367 cagtcattgtctcagaacaagcagtgagtgatccttcagacaggggacacaaaagt 426  
 1125 RRR 1066  
 427 t 427  
 1065 t 1065

## RESULT 2

US-09-276-531-29/C  
 Sequence 29, Application US/09276531  
 Patent No. 6183968

## GENERAL INFORMATION:

APPLICANT: Bandman, Olga  
 APPLICANT: Lal, Preeti  
 APPLICANT: Hillman, Jennifer L.  
 APPLICANT: Yue, Henry  
 APPLICANT: Reddy, Roopa  
 APPLICANT: Guebler, Karl J.  
 APPLICANT: Baughn, Mariah R.  
 TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING  
 NUMBER OF INVENTION: RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION  
 NUMBER OF SEQUENCES: 134  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
 STREET: 3174 PORTER DRIVE  
 CITY: PALO ALTO  
 STATE: CALIFORNIA  
 COUNTRY: USA  
 ZIP: 94304

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/276,531  
 FILING DATE: Herewith

CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/079,677  
 FILING DATE: March 27, 1998

CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:

NAME: Lynn E. Murry, Ph.D.  
 REGISTRATION NUMBER: 42,918  
 REFERENCE/DOCKET NUMBER: PA-0008 US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (650) 855-0555  
 TELEFAX: (650) 845-4166  
 INFORMATION FOR SEQ ID NO: 29:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 474 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 IMMEDIATE SOURCE:  
 LIBRARY: COLANTUT02  
 CLONE: 614640  
 US-09-276-531-29

Query Match 2.5%; Score 35.4; DB 4; Length 474;  
 Best Local Similarity 48.3%; Pred. No. 0.11;  
 Matches 99; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

339 tgataagtcagatcccaatgttaaacagttcatgtctccgaacacagtgaaagtgt 398  
 359 TAACATTTGAAGAACCAATACATAGACAGCGCTGTTCTTTCCCTCTTTCTTT 300  
 399 atccttgcaactggggaacacaagaatttatattcatttgaccctaaagtcttaac 458  
 299 TTAAAGAGAACTGGGACACCCCAATACATCTTAATTCATCTGACGTTCTCAG 240  
 459 aaatgaaagcgtctgt 518  
 239 ATATCAAGAAATGATGTCTTGTCTTTTGGCCAAAGAAAAGTGAAGCCTTAATTAC 180  
 519 ttatgtcctaaataccacagcaatt 543  
 179 TTTAATGACTCAGTCCACACAGATT 155

## RESULT 3

US-08-948-564-5/C  
 Sequence 5, Application US/08948564  
 Patent No. 6121512

## GENERAL INFORMATION:

APPLICANT: Siminszky, Balazs  
 APPLICANT: Dewey, Ralph E.  
 APPLICANT: Corbin, Frederick T.  
 TITLE OF INVENTION: No. 6121512el Cytochrome P-450 Constructs and  
 NUMBER OF INVENTION: Methods of Producing Herbicide-Resistant Transgenic Plants  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Virginia C. Bennett  
 STREET: PO Box 37428  
 CITY: Raleigh  
 STATE: No. 6121512th Carolina  
 COUNTRY: USA  
 ZIP: 27627

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/948,564  
 FILING DATE:

CLASSIFICATION: 800  
 ATTORNEY/AGENT INFORMATION:

NAME: Bennett, Virginia C.  
 REGISTRATION NUMBER: 37,092  
 REFERENCE/DOCKET NUMBER: 5051-409  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-854-1400  
 TELEFAX: 919-854-1401

INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1644 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 4..1542  
US-08-948-364-5

Query Match 2.4% Score 33; DB 4; Length 1644;  
Best Local Similarity 55.8%; Pred. No. 1.4;  
Matches 63; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 399 atccttgacatgggagacacaaagattattattcatttgactgaagttcttaac 458  
DB 756 AACCTGTTCTCGGACCTCCAGGCTTCTTCTTTGAGAAAAAGGGCTTAG 697  
QY 459 aaatggaagcgtctgtgtgtctgtagacagacaaaagaatcagctctca 511  
DB 696 AATGGAAGATGATCATCAATTCGGGTCACAAAGTATGATGAGAACACTCTTCA 644

## RESULT 4

US-08-318-492-3/C  
Sequence 3, Application US/08318492  
Patent No. 5552312

GENERAL INFORMATION:  
APPLICANT: Lim, Bing  
APPLICANT: Adra, Chaker N.  
TITLE OF INVENTION: RECOMBINANT HTM4 GENE, PROTEIN AND  
TITLE OF INVENTION: ASSAYS  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02173

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/318,492  
FILING DATE:

## CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:  
NAME: BROOK, David E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: BI94-03  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1661 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 97..741  
US-08-318-492-3

Query Match 2.3% Score 31.8; DB 1; Length 1661;  
Best Local Similarity 56.1%; Pred. No. 3.4;  
Matches 60; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 454 ttaacaatggaagcgtctgtgtctgtagacagacaaaagaatcagctcttaact 513  
DB 1352 TTATACATATCAGGCGCATGTGTGATTCACAGCTTCACTGAATAAATCCCTGAGCCTT 1293  
QY 514 gtacattatgtctcaaataccagctaattgctttaaagaagaga 560  
DB 1292 AACCTCCATCTCAAAATACAGTGAAATTTATTTATGATATATGA 1246

## RESULT 5

US-08-707-340-3/C  
Sequence 3, Application US/08707340  
Patent No. 5705615

GENERAL INFORMATION:  
APPLICANT: Lim, Bing  
APPLICANT: Adra, Chaker N.  
TITLE OF INVENTION: RECOMBINANT HTM4 GENE, PROTEIN AND  
TITLE OF INVENTION: ASSAYS  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02173

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/707,340  
FILING DATE: 03-SEP-1996  
CLASSIFICATION: 536

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/318,492  
FILING DATE: 06-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/675,648  
FILING DATE: 03-JUL-1996  
ATTORNEY/AGENT INFORMATION:

NAME: BROOK, David E.

REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: BI94-03A2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 3:

## SEQUENCE CHARACTERISTICS:

LENGTH: 1661 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:

NAME/KEY: CDS  
LOCATION: 97..738  
US-08-707-340-3

Query Match 2.3% Score 31.8; DB 1; Length 1661;  
Best Local Similarity 56.1%; Pred. No. 3.4;  
Matches 60; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 454 ttaacaatggaagcgtctgtgtctgtagacagacaaaagaatcagctcttaact 513  
DB 1352 TTATACATATCAGGCGCATGTGTGATTCACAGCTTCACTGAATAAATCCCTGAGCCTT 1293



Qy 514 gtacattatgtctcaaataccagctaattgcttttaagaagaaga 560  
| | | | | | | | | | | | | | | | | |  
Db 1292 AACCTCCCATCTCAAAATACAGTGAATAATTATTTTAATGATAATGA 1246

RESULT 6

```

US-08-994-578-3/c
Sequence 3, Application US/08994578
Patent No. 5972688
GENERAL INFORMATION:
APPLICANT: Lim, Bing
APPLICANT: Adia, Chaker N.
APPLICANT: Lellias, Jean-Michel
TITLE OF INVENTION: H7M4 METHODS OF TREATMENT AND ASSAYS
TITLE OF INVENTION: AGONISTS AND ANTAGONISTS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/994,578
FILING DATE: December 19, 1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/707,340
FILING DATE: 03-SEP-1996
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/675,648
FILING DATE: 03-JUL-1996
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/318,492
FILING DATE: 06-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: BIH94-0342Z
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781) 861-6240
TELEFAX: (781) 861-9540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1661 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 97..738
US-08-994-578-3

```

[illegible]

RESULT 7  
US-08-615-170-2

## GENERAL INFORMATION:

```

1  APPLICANT:  ORDAHL, Charles P.
2  APPLICANT:  AZAKIE, Anthony
3  APPLICANT:  MAR, Janet H.
4  APPLICANT:  FARRANCE, Iain K.G.
5  APPLICANT:  HALL, Deborah E.
6  APPLICANT:  STEWART, Alexandre F.R.
7  APPLICANT:  LARKIN, Sarah B.
8  TITLE OF INVENTION:  DTEF-1 ISOFORMS AND USES THEREOF
9  NUMBER OF SEQUENCES:  32
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE:  Townsend and Townsend Kourie and Crew
12 STREET:  Steuart Street Tower, One Market Plaza
13 CITY:  San Francisco
14 STATE:  California
15 COUNTRY:  US
16 ZIP:  94105-1493
17
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE:  Floppy disk
20 COMPUTER:  IBM PC compatible
21 OPERATING SYSTEM:  PC-DOS/MS-DOS
22 SOFTWARE:  PatentIn Release #1.0, Version #1.25
23 CURRENT APPLICATION DATA:
24 APPLICATION NUMBER:  US/08/615,170
25
26 FILING DATE:
27 CLASSIFICATION:  435
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER:  PCT/US95/01526
30 FILING DATE:  06-FEB-1995
31 CLASSIFICATION:  435
32 PRIOR APPLICATION DATA:
33 APPLICATION NUMBER:  US 08/191,493
34 FILING DATE:  04-FEB-1994
35 CLASSIFICATION:  435
36 ATTORNEY/AGENT INFORMATION:
37 NAME:  Heslin, James M.
38 REGISTRATION NUMBER:  29,541
39 REFERENCE/DOCKET NUMBER:  23070-U-053120
40 TELECOMMUNICATION INFORMATION:
41 TELEPHONE:  (415) 326-2400
42 TELEFAX:  (415) 326-2422
43 INFORMATION FOR SEQ ID NO:  2:
44 SEQUENCE CHARACTERISTICS:
45 LENGTH:  1627 base pairs
46 TYPE:  nucleic acid
47 STRANDEDNESS:  single
48 TOPOLOGY:  linear
49 MOLECULE TYPE:  cdna
50 FEATURE:
51 NAME/KEY:  CDS
52 LOCATION:  185..1480
53 OTHER INFORMATION:  /product= "TEF-1A"
54
55 US-08-615-170-2

```

Query Match	2.28;	Score 31.4;	DB 1;	Length 1627;
Best Local Similarity	48.6%;	Pred. No. 4.6;		
Matches 86;	Conservative 0;	Mismatches 91;	Indels 0;	Gaps 0
Qy 392	aaggtgatactctggcaactgagggaacacaaagtatttattatctatcttgacccaagt	451		
Db 1119	ATGGTGTTCACGTCATATATGAGAGCCCAAAATATGTGCATTACTCTGTCACCTAAAG	1178		
Qy 452	tcttaacaagaatgaagcgtgtctgtgtttcttggagacagacagaagaatcagctcttca	511		
Db 1179	TGTGTTCCTTTGGAAAGCAGAGTGATGAGAAAGTGGAGACAGAGTATGCACACTATGAAA	1238		
Qy 512	ctgtacattatgtcttcaaataccacagtaattgtctttaaagaagaagacataact	568		

Db 1239 ATGACACTATGCTATGCAATGATGCTCTCTCTGTAATACATGATTAACCT 1295

RESULT 8

US-08-615-170-4

Sequence 4, Application US/08615170

Patent No. 5776776

GENERAL INFORMATION:

APPLICANT: ORDAHL, Charles P.

APPLICANT: AZAKIE, Anthony

APPLICANT: FARANCE, Iain K.G.

APPLICANT: HALL, Deborah E.

APPLICANT: STEWART, Alexandre F.R.

APPLICANT: LARKIN, Sarah B.

TITLE OF INVENTION: DTEF-1 ISOFORMS AND USES THEREOF

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESS: Townsend and Townsend Kourie and Crew

STREET: Stewart Street Tower, One Market Plaza

CITY: San Francisco

STATE: California

COUNTRY: US

ZIP: 94105-1493

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/615,170

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/01526

FILING DATE: 06-FEB-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/191,493

FILING DATE: 04-FEB-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Heslin, James M.

REGISTRATION NUMBER: 29,541

REFERENCE/DOCKET NUMBER: 2307U-053120

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 326-2400

TELEFAX: (415) 326-2422

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1666 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 185..1519

OTHER INFORMATION: /product="TEF-1B"

FEATURE:

NAME/KEY: misc\_feature

LOCATION: 536..571

OTHER INFORMATION: /product="KTQYDKYFSSEK"

US-08-615-170-4

Query Match 2.2%; Score 31.4; DB 1; Length 1666;  
Best Local Similarity 48.6%; Pred. No. 4.6;  
Matches 86; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

392 aaggtatccttgcaactgggaacacaaagatttattattcattgactcaagt 451

Db 1158 ATGCTGTTCCACTCAATATGAGAGCCGCAAAATATGTCATTACTGTTCCACTAAG 1217

QY 452 tcttaacaatggaagcgtctgttcttggaagcagcaagaataagatctctca 511

Db 1218 TGTGTTCTTTGGAAAGCAGGTGTGGAGAAAGTGAAGATGACATGATGATGATA 1277

QY 512 ctgtaattatgtctcaaataccagctaattgcttttaagaagaagacatact 568

Db 1278 ATGACACTATGCTATGCAATGATGCTCTCTCTGTAATACATGATTAACCT 1334

RESULT 9

US-08-781-891-209

Sequence 209, Application US/08781891

Patent No. 6090620

GENERAL INFORMATION:

APPLICANT: Fu, Ying-Hui

APPLICANT: Yu, Chang-Eu

APPLICANT: Oshima, Junko

APPLICANT: Mulligan, John T.

APPLICANT: Schellenberg, Gerald D.

TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO

NUMBER OF SEQUENCES: 209

CORRESPONDENCE ADDRESS:

ADDRESS: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/781,891

FILING DATE: 27-DEC-1996

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: NO. 6090620tendburg Ph.D., Carol

REGISTRATION NUMBER: 39,317

REFERENCE/DOCKET NUMBER: 240052.419

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 682-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 209:

SEQUENCE CHARACTERISTICS:

LENGTH: 51259 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-781-891-209

Query Match 2.2%; Score 31.4; DB 3; Length 51259;  
Best Local Similarity 50.3%; Pred. No. 40;  
Matches 77; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 537 gctaattgctttaagaagaagacatactatgacgagccagagaacatcatgag 596

Db 8997 GATATATGATATAGAGAAAGTGAAGATTCCCACTCAAGAGCCAGTAAATATCTTCA 9056

QY 597 cacaagtaccgggacctgctacgtacgtacgtacgtacgtacgtacgtacgtacgt 656

Db 9057 CAAATATATAGAAAGAACTTCATATACCTTAAGAAAGCATGTTCATGAACATACAGA 9116

QY 657 agctgtcaagccaagaagaataatgccaagaa 689

Db 9117 AACCTTCAGAACTCCAAATATAGACTGACACAGAA 9149

RESULT 10  
US-08-737-524B-26  
Sequence 26, Application US/08737524B  
Patent No. 5912414  
GENERAL INFORMATION:  
APPLICANT: CARL SAVERIO FALCO  
APPLICANT: DOMINICK ANTHONY GUIDA, JR.  
APPLICANT: MARY ELIZABETH HARNETT LOCKE  
TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS, CHIMERIC  
TITLE OF INVENTION: GENES AND METHODS FOR INCREASING  
TITLE OF INVENTION: THE METHIONINE CONTENT OF THE SEEDS  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
STREET: 1007 MARKET STREET  
CITY: WILMINGTON  
STATE: DELAWARE  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 19898  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.50 INCH  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MICROSOFT WINDOWS 95  
SOFTWARE: MICROSOFT WORD FOR WINDOWS 95 (7.0)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/737,524B  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: LYNNE M. CHRISTENSEN  
REGISTRATION NUMBER: 30,971  
REFERENCE/DOCKET NUMBER: BB-1059-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 302-992-5481  
TELEFAX: 302-773-0164  
TELEX: 833420  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3639 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-737-524B-26  
Query Match 2.2%; Score 31; DB 2; Length 3639;  
Best Local Similarity 50.3%; Pred. No. 10;  
Matches 76; Conservative 0; Mismatches 75; Indels 0; Gaps 0;  
QY 426 ttattatattcatttgacctcaagttcttaacaatgaagcgtgtgtgttcttga 485  
DB 3200 tctgagatcacaacacgcatctctgtttacatgaactagattctgtcttga 3259  
QY 486 gacgacgaaagaaatgaagctcttctacatattgtctcaaataccagctaattgc 545  
DB 3260 AAGGAAGAATAGGATATGTTCTGTGCACTGCATATATATCATATTAATCGGAGC 3319  
QY 546 tttaagaagaagacatactatctgtc 576  
DB 3320 TTTATGTCACAGACTCACAGCCAGGCTACG 3350  
RESULT 11  
US-08-147-812-4/C  
Sequence 4, Application US/08147812  
Patent No. 5766909  
GENERAL INFORMATION:  
APPLICANT: Xie, Qiao-wen  
APPLICANT: Nathan, Carl F.  
APPLICANT: Mumford, Jimmy A.  
APPLICANT: Calaycay, Jimmy Ramos

TITLE OF INVENTION: DNA Encoding Inducible Nitric Oxide Synthase  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merck & Co., Inc.  
STREET: 126 East Lincoln Avenue  
CITY: Rahway  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07065  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: Macintosh Centris650  
OPERATING SYSTEM: Macintosh 7.0.1  
SOFTWARE: Microsoft Word 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/147,812  
FILING DATE: No. 5766909 Available  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/841,641  
FILING DATE: 02-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wallen, John W III  
REGISTRATION NUMBER: 35,403  
REFERENCE/DOCKET NUMBER: 186581A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 594-3905  
TELEFAX: (908) 594-4720  
TELEX: 138825  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4041 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-147-812-4  
Query Match 2.2%; Score 30.8; DB 1; Length 4041;  
Best Local Similarity 50.7%; Pred. No. 13;  
Matches 74; Conservative 0; Mismatches 72; Indels 0; Gaps 0;  
QY 613 ctgttactgaacctgaagaagtggtgttcttccacacaaagctgtcaagcaaca 672  
DB 2166 ctgttgatgtatgtatgacgaagccagacgaagctgacgtatcttgagccaa 2107  
QY 673 agaatatgcccagaagagtggtgttgaagtcgaagaggaaggtcttgaacac 732  
DB 2106 CACAGCATCTGCAAGCTGTGTGTTCTCTAAGCATGAACAGATTTCTTCAGAGT 2047  
QY 733 attacatctctacacagccacat 758  
DB 2046 CTGCCCATCTGTGGACAGCTTCAT 2021  
RESULT 12  
US-09-123-708-1/C  
Sequence 1, Application US/09123708  
Patent No. 6146887  
GENERAL INFORMATION:  
APPLICANT: SCHRAEDER, Juergen  
APPLICANT: GORDEKE, Axel  
TITLE OF INVENTION: DNA EXPRESSION VECTORS FOR USE IN GENE THERAPEUTIC  
TITLE OF INVENTION: TREATMENT OF VASCULAR DISORDERS  
FILE REFERENCE: 511169-2003  
CURRENT APPLICATION NUMBER: US/09/123,708  
CURRENT FILING DATE: 1998-07-28  
EARLIER APPLICATION NUMBER: 08/553,503  
EARLIER FILING DATE: 1996-03-01  
EARLIER APPLICATION NUMBER: P4411402.8  
EARLIER FILING DATE: 1994-03-31  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1  
LENGTH: 4110  
TYPE: DNA  
ORGANISM: Cytomegalovirus  
US-09-123-708-1

Query Match  
Best Local Similarity 2.2%; Score 30.8; DB 4; Length 4110;  
Matches 74; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 613 ctgttcactgaactcaggaagagtggtcttccaacaaagctgtcaagccaaca 672  
DB 2102 CTGTCGATGTCATGAGCAAGGCCGAGACTGAGGTCATGTCGAGCCAAAGCCCAA 2043  
QY 673 agaatatgcccaagaagtggttgatgtgacgaaggaagggcttccttgacaac 732  
DB 2042 CACAGCATACCTGAAGGTGTGTGAGTCTCTAAGCATGAAACAGAGATTCTTCAGAGT 1983  
QY 733 attaccatctcaccacagccacacat 758  
DB 1982 CTGCCCATTCCTGGGACAGTCTCCAT 1957

RESULT 13  
US-09-123-624-1/c  
Sequence 1, Application US/09123624  
Patent No. 6149936  
GENERAL INFORMATION:  
APPLICANT: SCHRAEDER, Jurgen  
APPLICANT: GODKE, Axel  
TITLE OF INVENTION: DNA EXPRESSION VECTORS FOR USE IN THE GENE THERAPEUTIC  
FILE REFERENCE: 51169-2004  
CURRENT FILING DATE: 1998-07-28  
CURRENT APPLICATION NUMBER: 08/553,503  
PRIOR FILING DATE: 1996-03-01  
PRIOR APPLICATION NUMBER: 4411402.8  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4110  
TYPE: DNA  
ORGANISM: Mus musculus  
US-09-123-624-1

Query Match  
Best Local Similarity 2.2%; Score 30.8; DB 4; Length 4110;  
Matches 74; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 613 ctgttcactgaactcaggaagagtggtcttccaacaaagctgtcaagccaaca 672  
DB 2102 CTGTCGATGTCATGAGCAAGGCCGAGACTGAGGTCATGTCGAGCCAAAGCCCAA 2043  
QY 673 agaatatgcccaagaagtggttgatgtgacgaaggaagggcttccttgacaac 732  
DB 2042 CACAGCATACCTGAAGGTGTGTGAGTCTCTAAGCATGAAACAGAGATTCTTCAGAGT 1983  
QY 733 attaccatctcaccacagccacacat 758  
DB 1982 CTGCCCATTCCTGGGACAGTCTCCAT 1957

RESULT 14  
US-08-147-812-6/c  
Sequence 6, Application US/08147812  
Patent No. 5766909  
GENERAL INFORMATION:  
APPLICANT: Xie, Qiao-wen  
APPLICANT: Nathan, Carl F.

APPLICANT: Mumford, Richard A.  
APPLICANT: Calaycay, Jimmy Ramos  
TITLE OF INVENTION: DNA Encoding Inducible Nitric Oxide Synthase  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESS: Merck & Co., Inc.  
STREET: 126 East Lincoln Avenue  
CITY: Rahway  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07065

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: Macintosh Centris650  
OPERATING SYSTEM: Macintosh 7.0.1  
SOFTWARE: Microsoft Word 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/147,812  
FILING DATE: No. 5766909 Available  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/841,641  
FILING DATE: 02-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wallen, John W III  
REGISTRATION NUMBER: 35,403  
REFERENCE/DOCKET NUMBER: 186581A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 594-3905  
TELEFAX: (908) 594-4720  
TELEX: 138825

INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4165 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-147-812-6

Query Match  
Best Local Similarity 2.2%; Score 30.8; DB 1; Length 4165;  
Matches 74; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 613 ctgttcactgaactcaggaagagtggtcttccaacaaagctgtcaagccaaca 672  
DB 2166 CTGTCGATGTCATGAGCAAGGCCGAGACTGAGGTCATGTCGAGCCAAAGCCCAA 2107  
QY 673 agaatatgcccaagaagtggttgatgtgacgaaggaagggcttccttgacaac 732  
DB 2106 CACAGCATACCTGAAGGTGTGTGAGTCTCTAAGCATGAAACAGAGATTCTTCAGAGT 2047  
QY 733 attaccatctcaccacagccacacat 758  
DB 2046 CTGCCCATTCCTGGGACAGTCTCCAT 2021

RESULT 15  
US-08-701-582D-1  
Sequence 1, Application US/08701582D  
Patent No. 6017755  
GENERAL INFORMATION:  
APPLICANT: WRANA, Jeffrey  
APPLICANT: ATTISANO, Lilliana  
APPLICANT: SCHERER, Stephen W.  
TITLE OF INVENTION: MADR2 TUMOR SUPPRESSOR GENE  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESS: Burns, Doane, Swecker & Mathis  
STREET: P O Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: U.S.A.

Fri May 25 10:28:26 2001

us-09-403-269-12\_copy\_1\_1404.rni

ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/701,582D  
FILING DATE: 22-AUG-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: REA, Teresa Stanek  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 024916-004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1605 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-701-582D-1

Query Match 2.28; Score 30.6; DB 3; Length 1605;  
Best Local Similarity 49.18; Pred. No. 8.3; Mismatches 84; Indels 0; Gaps 0;  
Matches 81; Conservative 0; Mismatches 0;  
QY 1024 atgaataacatgactgcatgaagaatcccaactacacactagctcttltgtttaaat 1083  
DB 229 atcaccactcaaaactgtaataatgtgtacacataccacacactgtctgaaatt 288  
QY 1084 ggccttaagtattcttattatgggctgatacttaagaagaactgcaggagaaactc 1143  
DB 289 tggggactgtagtaccacaaatgacatgacatgagtggtacacacagccctttacagcttc 348  
QY 1144 gggaaagaagcgaggtccctgtagagcgttgacatgaaatccctt 1188  
DB 349 tctgaacaaacacagctctcttgatgtgcgtccacagtatcccat 393

Search completed: May 25, 2001, 03:55:37  
Job time: 5593 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2001, 01:57:43 ; Search time 1044.13 Seconds  
(without alignments)  
11747.083 Million cell updates/sec

Title: US-09-403-269-12\_COPY\_1\_1404  
Sequence: 1404  
1 tccaagcgaattctcatag.....gcagccgagcaagacaaac 1404

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 9623517 segs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1:	gb_est1:*
2:	gb_est2:*
3:	gb_est3:*
4:	gb_est4:*
5:	gb_est5:*
6:	gb_est6:*
7:	gb_est7:*
8:	gb_est8:*
9:	gb_est9:*
10:	gb_est10:*
11:	gb_est11:*
12:	gb_est12:*
13:	gb_est13:*
14:	gb_est14:*
15:	gb_est15:*
16:	gb_est16:*
17:	gb_est17:*
18:	gb_est18:*
19:	gb_est19:*
20:	gb_est20:*
21:	gb_est21:*
22:	gb_est22:*
23:	gb_est23:*
24:	gb_est24:*
25:	gb_est25:*
26:	gb_est26:*
27:	gb_est27:*
28:	gb_est28:*
29:	gb_est29:*
30:	gb_est30:*
31:	gb_est31:*
32:	gb_est32:*
33:	gb_est33:*
34:	gb_est34:*
35:	gb_est35:*
36:	gb_est36:*
37:	gb_est37:*
38:	gb_est38:*
39:	gb_est39:*
40:	gb_est40:*
41:	gb_est41:*
42:	gb_est42:*
43:	gb_est43:*

44:	em_esthum10:*
45:	em_esthum11:*
46:	em_esthum12:*
47:	em_esthum13:*
48:	em_esthum14:*
49:	em_esthum15:*
50:	em_esthum16:*
51:	em_esthum17:*
52:	em_esthum18:*
53:	em_esthum19:*
54:	em_esthum20:*
55:	em_esthum21:*
56:	em_esthum22:*
57:	em_esthum23:*
58:	em_esthum24:*
59:	em_esthum25:*
60:	em_esthum26:*
61:	em_esthum27:*
62:	em_esthum28:*
63:	em_esthum29:*
64:	em_esthum30:*
65:	em_esthum31:*
66:	em_esthum32:*
67:	em_esthum33:*
68:	em_esthum34:*
69:	em_esthum35:*
70:	em_esthum36:*
71:	em_esthum37:*
72:	em_esthum38:*
73:	em_esthum39:*
74:	em_esthum40:*
75:	em_esthum41:*
76:	em_esthum42:*
77:	em_esthum43:*
78:	em_esthum44:*
79:	em_esthum45:*
80:	em_esthum46:*
81:	em_esthum47:*
82:	em_esthum48:*
83:	em_esthum49:*
84:	em_esthum50:*
85:	em_esthum51:*
86:	em_esthum52:*
87:	em_esthum53:*
88:	em_esthum54:*
89:	em_esthum55:*
90:	em_esthum56:*
91:	em_esthum57:*
92:	em_esthum58:*
93:	em_esthum59:*
94:	em_esthum60:*
95:	em_esthum61:*
96:	em_esthum62:*
97:	em_esthum63:*
98:	em_esthum64:*
99:	em_esthum65:*
100:	em_esthum66:*
101:	em_esthum67:*
102:	em_esthum68:*
103:	em_esthum69:*
104:	em_esthum70:*
105:	em_esthum71:*
106:	em_esthum72:*
107:	em_esthum73:*
108:	em_esthum74:*
109:	em_esthum75:*
110:	em_esthum76:*
111:	em_esthum77:*
112:	em_esthum78:*
113:	em_esthum79:*
114:	em_esthum80:*
115:	em_esthum81:*
116:	em_esthum82:*

117: gb\_est48:\*  
 118: gb\_est49:\*  
 119: gb\_est50:\*  
 120: gb\_est51:\*  
 121: gb\_est52:\*  
 122: gb\_est53:\*  
 123: gb\_est54:\*  
 124: gb\_est55:\*  
 125: gb\_est56:\*  
 126: gb\_est57:\*  
 127: gb\_est58:\*  
 128: gb\_est59:\*  
 129: gb\_est60:\*  
 130: gb\_est61:\*  
 131: gb\_est62:\*  
 132: gb\_est63:\*  
 133: gb\_est64:\*  
 134: gb\_est65:\*  
 135: gb\_est66:\*  
 136: gb\_est67:\*  
 137: gb\_est68:\*  
 138: gb\_est69:\*  
 139: gb\_est70:\*  
 140: gb\_est71:\*  
 141: gb\_est72:\*  
 142: gb\_est73:\*  
 143: gb\_est74:\*  
 144: gb\_est75:\*  
 145: gb\_est76:\*  
 146: gb\_est77:\*  
 147: gb\_est78:\*  
 148: gb\_est79:\*  
 149: gb\_est80:\*  
 150: gb\_est81:\*  
 151: gb\_est82:\*  
 152: gb\_est83:\*  
 153: gb\_est84:\*  
 154: gb\_est85:\*  
 155: gb\_est86:\*  
 156: gb\_est87:\*  
 157: gb\_est88:\*  
 158: gb\_est89:\*  
 159: gb\_est90:\*  
 160: gb\_est91:\*  
 161: gb\_est92:\*  
 162: gb\_est93:\*  
 163: gb\_est94:\*  
 164: gb\_est95:\*  
 165: gb\_est96:\*  
 166: gb\_est97:\*  
 167: gb\_est98:\*  
 168: gb\_est99:\*  
 169: gb\_est100:\*  
 170: gb\_est101:\*  
 171: gb\_est102:\*  
 172: gb\_est103:\*  
 173: gb\_est104:\*  
 174: gb\_est105:\*  
 175: gb\_est106:\*  
 176: gb\_est107:\*  
 177: gb\_est108:\*  
 178: gb\_est109:\*  
 179: gb\_est110:\*  
 180: gb\_est111:\*  
 181: gb\_est112:\*  
 182: gb\_est113:\*  
 183: gb\_est114:\*  
 184: gb\_est115:\*  
 185: gb\_est116:\*  
 186: gb\_est117:\*  
 187: gb\_est118:\*  
 188: gb\_est119:\*  
 189: gb\_est120:\*

190: em\_gss\_pln1:\*  
 191: em\_gss\_pln2:\*  
 192: em\_gss\_pro:\*  
 193: em\_gss\_rod1:\*  
 194: em\_gss\_rod2:\*  
 195: em\_gss\_rod3:\*  
 196: em\_gss\_rod4:\*  
 197: em\_gss\_rod5:\*  
 198: em\_gss\_vrt1:\*  
 199: em\_gss\_vrt2:\*  
 200: em\_gss\_vrt3:\*  
 201: gb\_gss1:\*  
 202: gb\_gss2:\*  
 203: gb\_gss3:\*  
 204: gb\_gss4:\*  
 205: gb\_gss5:\*  
 206: gb\_gss6:\*  
 207: gb\_gss7:\*  
 208: gb\_gss8:\*  
 209: gb\_gss9:\*  
 210: gb\_gss10:\*  
 211: gb\_gss11:\*  
 212: gb\_gss12:\*  
 213: gb\_gss13:\*  
 214: gb\_gss14:\*  
 215: gb\_gss15:\*  
 216: gb\_gss16:\*  
 217: gb\_gss17:\*  
 218: gb\_gss18:\*  
 219: gb\_gss19:\*  
 220: gb\_gss20:\*  
 221: gb\_gss21:\*  
 222: gb\_gss22:\*  
 223: gb\_gss23:\*  
 224: gb\_gss24:\*  
 225: gb\_gss25:\*  
 226: gb\_gss26:\*  
 227: gb\_gss27:\*  
 228: gb\_gss28:\*  
 229: gb\_gss29:\*  
 230: gb\_gss30:\*  
 231: gb\_gss31:\*  
 232: gb\_gss32:\*  
 233: gb\_gss33:\*  
 234: gb\_gss34:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	599.6	42.7	698	103	AI860635 w104a09.x
C 2	586.2	41.8	776	107	AU139373 AU139373
C 3	521.1	37.1	653	114	AW83385 PM3-HT034
C 4	450.4	32.1	559	21	AI521253 t066109.x
C 5	401.8	28.6	477	114	AW83386 PM3-HT034
C 6	364.2	25.9	426	7	AA435342 v049909.s
C 7	346.6	24.7	455	12	AA794925 v148904.s
C 8	331.6	23.6	866	230	CNSOIWIT
C 9	304.2	21.7	358	20	AI467374 v049909.x
C 10	303	21.6	434	163	BE170776 BE170776
C 11	285.4	20.3	566	143	BE080558 BE080558
C 12	285.4	20.3	590	104	AI981391 pat.pk005
C 13	195	13.9	361	140	BE861166 UI-M-AW0-
C 14	164.6	11.7	251	160	BB579985 BB579985
C 15	143.6	10.2	604	31	AV672171 AV672171
C 16	139.4	9.9	203	117	AW605325 OV3-DT004
C 17	129.2	9.2	663	31	AV676110 AV676110
C 18	104.8	7.5	802	230	CNS02E30
C 19					AI193221 Tetradon



## ALIGNMENTS

Query Match	Local Similarity	Score	DB	Length
Best Local	91.18	599.6	103	698
Matches	633	Conservative	0	Mismatches 62; Indels 0; Gaps 0
QY	359	atglttaaacagcttcattgctccgaagaaacagctgaagatgtracctctgcacacttggggaaaca	418	
Db	697	ATGTCAAAACAGTTAATATGACCAAAAAACAGTGAAGGTGATTTTTCGAACTGGAAACAC	638	
QY	419	caaaaggtttatattatttcatttgacctcaagttcttcaacaaatggaagcgctgctgtg	478	
Db	637	AAAMGATTTTATTTATTCCTCATTTTGACCCTCANGTCTTGTGCAATATGNAATGATGTGCTGGT	578	
QY	479	ttcttggagacgaagaagaalacagctcttcactgttacctgtatcatatgtcctaataatccggc	538	
Db	577	TTCCTAGAGACCAACAAAAAAGATCAGCTCTTCACATATACATTAATATGCTCAACAC	518	
QY	539	taatttccttttaagaagaagacatactactatgcaatcgggcccgacaacatcatgtagaca	598	
Db	517	TAAATTCCTTTTAAAGAAAAAGATATATATCTATGCAATTTGGCCCAAGAACTTCATGAGCA	458	
QY	599	cagttacccgggacccgctgcactgctacgtccctgaagaagagtggtgtcttccacacaaag	658	
Db	457	CAGTTACCCGGGACCTGGTCACGTGACCTCAGSAAAGAGATGGGTCTTTCAACACCAAAAG	398	
QY	659	ctgtcaagccacaagaagataatgccccagaaggtggttaagttgatgtcgaaagggaaag	718	
Db	397	CTGTCAAGCCAAACCAAAATATGCCCCAAGAAAGTGTTGATGATTCAAAGAGTAGG	338	
QY	719	gcttccttgaacacattaccatctctacacagagccacatgctgcctctctctgtgcga	778	
Db	337	GATTCCTCGACAAACATPACATCTCTACACAGCCACATGGCTGCATTTTTCGTGCTGA	278	
QY	779	gtagctctctgtgttgaggaaacaagatgtgaaagcgcgtgcgcgattatgltgaacogta	838	
Db	277	GTGATTTGGCTAGTAAGGAACACAGAGTGAAGAAAGGTGGCCCAATATATGCTGACCCGTA	218	
QY	839	agtttaggggaaggtctcaagctctttagagccaaggtgtgtactccgcaatgtgcccgaagggc	898	
Db	217	AGTTAAGGGGAAGGGTTCAAGCTTTTAGAGCCAGAGTGTATTTTGCCATGGCCCAAGGGC	158	
QY	899	aagccattctacatctagtctaggggcctatctcttaacaaaagaccatatatctcctaatt	958	
Db	157	AAGCATTTTCTACATTAATCTCAGGGCCTATCTGTTAACAAAAAGACCAATATATTCCTCAATT	98	
QY	959	cagctttaaaggcaacagcccccctacaaagtctctgtcagaagcagcatgtgaatcaagctg	1018	
Db	97	CAGCTTTAAAGGCAACAGCCCTTATTAAGTTTTCATCTGAGCAGACATGAGGTTAAAGCTG	38	
QY	1019	tgattatgaataacaactgactggtgataagaatatatcc	1055	
Db	37	TGTTATGATTAACATGACTGCTGATTAACAAATATCC	1	

REFERENCE AUTHORS	TITLE JOURNAL COMMENT
1 (Passes 1 to 776) Oca,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.	HRI human cDNA project Unpublished (2000) Contact: Takao Isogai Genomics Laboratory

BASE COUNT	248 a	151 c	177 g	197 t	3 others
ORIGIN					

	Query Match	Similarity	91.8%	Score	586.2	DB	107	Length	776
	Best Local	Similarity	92.7%	Pred. No.	1.4e-162				
	Matches	659	Conservative	0	Mismatches	45	Indels	7	Gaps
Qy	13	lctcaatgatacttccaagatctatgtcagacagagagcccccattaccctgagtggtctt	72						
Db	53	TCTCATACTATTTCACAAAGCTATATGACACAGAGGCCCTTATACCCCGATGTGTCTTT	112						
Qy	73	atgtcccttgaagctacaaatgtcgaagtcgcgaagaagatcaagltgcataagtggagtt	132						
Db	113	ATGCTTTTGAAGGCTCAATGTGGAAGTCCGAGACAGAGTCACAGTCATATAGTGGGTT	172						
Qy	133	gaaggtgtaaccttatactaacagagtgggggagcccaaggtatcttcaccacacagatt	192						
Db	173	GAAAGTGTGCCATTATCTACACAAATGGGACCTCAAGGCTATTTCATACCAATCCAGATT	232						
Qy	193	gcacagatctgggttaagtcactacacagaatctactgaaaaacccctcatataag	252						
Db	233	GCACAGTATGGATTATAGTCATTACGACAAATCTAATCGAAGACCTCTCATTAAAG	292						
Qy	253	gatatgtgaagaacgggaagacagagacgaagaacagacagccatgactgcagcttgccc	312						
Db	293	GTAATATGAACACCGAAGAGACAGACAGAAAA---CAAGCTATATGACTGTGACTGGCCA	349						
Qy	313	aaggagctgcttatavggctagtgtggctgataagtaagtaacagattcaacaaatctaaagttc	372						
Db	350	AAGGGCTGCTTTATATGGCAATGTGGCTGTATAGTCTAGATTCCACAAATGCAACAGTTT	409						
Qy	373	atgtgcccgaaacacagtgaaagtgtatcccttgtaaccttgggagacacaaagaatttatt	432						
Db	410	ATTGCACCGAAGAACCAAGTAGAGGTATTCCTTGCACATGGGAACCAAAAGATTTTAT	469						
Qy	433	atttcattgacctcaagttctttaaacaatbgaagcgtgtcttggttct-agaagacgac	491						
Db	470	ATTTCATTTTGACCTCAAGTTCTTGACAAATGGAAGTGTGTGCTGTCTTAGAAGCAC	529						
Qy	492	agaaagaatcagctcttcacgttaccatcatgtctcaataaccacagtaattgctttaa	551						
Db	530	AGAAAGAAATACAGCTCTTCACTATCATTTATGTCTCAAAATGCTCAGCTAATTCCTTTAA	589						
Qy	552	agaagaagacatactaat-aggcatcggggcccaagaacatactgagacaagttaccggg	610						
Db	590	AGAAAGAAATATCTATGTGGCTATTTGGGCCACAGAACTCATGTGAGACAGATTACAGGG	649						

[illegible]

					Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asampson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL <a href="http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM3&amp;lc2=PM3-HM0347-091159-001-d06&amp;ct3=1999-11-09&amp;cl4=1">http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM3&amp;lc2=PM3-HM0347-091159-001-d06&amp;ct3=1999-11-09&amp;cl4=1</a> ) Seq primer: puc 18 forward High quality sequence start: 38 High quality sequence stop: 567. Location/Qualifiers 1..653 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="HM0347" /dev_stage="Adult" /note="Organ: head-neck; Vector: puc18; Site.1: SmaI; Site.2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) files into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."			
BASE COUNT	215	a	131	c	143	g	164	t
ORIGIN								
Query Match			37.1%		Score 521;	DB 114;	Length 653;	
Best Local Similarity			92.1%		Pred. No. 2.9e+143;			
Matches 572;		Conservative	0;	Mismatches	45;	Indels	4;	Gaps 2.
Dy	183	aatccagatgcacagatacgtatgaagtcaactcacgacagaatctaactcgaaanaaccgcc	242					
	8	AATCCAAATTGCACAGTATGATTGAATTAATCTCTTTTCACGAAATAATCTAATCTAGAAAACCTCC	67					
Dy	243	tcatatgagtgatatgaacaacagacagagaacagacagacaaaacagcaaggccaagtactg	302					
	68	TCAATATGAGGTATATATGAACAACAGCAGAAGAACAGACAAAA---CAAAGCTAATGACTG	124					
Dy	303	gactgtgcccaagggtcgtcttatgtactgtactgtggctgcgataaagtcgaattcaccaatgt	362					
	125	GACTGTGCCAAAGGGCTGCTTTATGGCGCAATGTGGTGATTAAGTCTAGATTCCCAATGT	184					
Dy	363	taaacagttaattgtctccagaaacacagtcgaagtgtatcctctgcaactggyggaacaaaa	422					

```

Db 185 CAAACAGTTATTTGACACAGAAACAGTGAAGGTATCTTCCACTGGAAACACAA 244
Qy 423 agattatattatcattgaccccaagttcttaacaaaggaagcgtctgtgttct 482
Db 245 AGATTATATATTCATTGACCTCAAGTCTTCCAAAGAAAGAGTGTCTCGTGTCT 304
Qy 483 ggaagacagcaaaagaatagctcttcaactgacatattatgtctcaataccagctat 542
Db 305 AGAGACCCAGAAAGAAAGTACCTTTCACATATATATATATATATATATATATAT 364
Qy 543 tgcctttaaagaagagacatatactatgcatcgagcccaagaacatcttgagcagct 602
Db 365 TGGCTTTAAAGAAAGATATATATATATATATATATATATATATATATATATATAT 424
Qy 603 taccgggagctgtgacactgacacagaaagagtggtgttcttccacacaaagcgtc 662
Db 425 TACACAGGAGCTGTGCTACGACCTCAGAAAGAGTGGGTCTTTTAAACACAAAGCTGT 484
Qy 663 caagccacaagaataatgccaagaagtggtgttcttgcgaaggaaggaagcgtc 722
Db 485 CAAGCCCAACCAATATATATATATATATATATATATATATATATATATATATATAT 544
Qy 723 ccttgacaacatatacctctctacacagccacacagctgc-cctctgtctcagtg 781
Db 545 CCTCGACACATTTACCATCTCTACACAGCCACATGTTGCTATTTTCTCTCTAGTG 604
Qy 782 actggtgtgtgaggaagcag 802
Db 605 ATTGGCTAGTAAAGAACACAG 625

RESULT 4
LOCUS A1521253 559 bp mRNA EST 12-MAY-1999
DEFINITION to66h09.x1 NCI CGAP Gas4 Homo sapiens cDNA clone IMAGE:2183297 3'
sequence.
ACCESSION A1521253
VERSION A1521253.1 GI:4435388
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 559)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILM at:
www.bio.livnl.gov/brp/image/image.html
Insert Length: 1602 Std Error: 0.00
Seq Primer: -40UP from Gibco
High quality sequence stop: 376
POLYA=No.

FEATURES
Source Location/Qualifiers
1..559
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2183297"
/clone_lib="NCI CGAP Gas4"
/tissue_type="poorly differentiated adenocarcinoma with
signed ring cell features"

```

```

/lab_host="DH10B"
/notes="Organ: Stomach; Vector: pCMV-Sport6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.69 kb. Life Technologies catalog #:
11549-011"
BASE COUNT 145 a 130 c 113 g 169 t 2 others
ORIGIN

```

```

Query Match 32.1%; Score 450.4; DB 21; Length 559;
Best Local Similarity 90.9%; Pred. No. 2.4e-122;
Matches 511; Conservative 0; Mismatches 48; Indels 3; Gaps 3;

```

```

Qy 490 acagaagaatcagctcttcaactgacatattatgtctcaataaccagctaatgtctt 549
Db 559 ACAGAAAGATCAGCTTCTTCACTATATATATATATATATATATATATATATATATAT 500
Qy 550 aagaagaagacatactatgcatcggtcccaagaacatcatgagacagcttaccgg 609
Db 499 -AAGAAAGATATATATATATATATATATATATATATATATATATATATATATAT 441
Qy 610 gacctgtcactgacactcaggaagagtggtgttcttccacacaaagctgtcaagca 669
Db 440 GACTTGCTACTG-CCTCAGGAAAGAGTGGGTCTTTTCAACACAAAGCTGTCAAGCCA 382
Qy 670 acagaataatgcccagaagtggtgttgaattgttggaaggaagggcttccctgac 729
Db 381 ACCAANATATATATATATATATATATATATATATATATATATATATATATATATAT 322
Qy 730 aacattacatctctacacagccacagctgtcctctctgtcctcagtgacgtgtc 789
Db 321 -ACATTACCATCTCTACACAGCCACAGCTGTGCTATTTTGTCTGTATGATGAGCTA 263
Qy 790 gtgaggaaccagagatgagaagagcgtgtgacgtatgttgagccctgaagtgaggga 849
Db 262 GTAAGGAACAGAGATGAGAAAGGTGCTGCGCAATATATGCTGACCCGTATAGGGGAA 203
Qy 850 ggctcagctctttagagccagagtggtgttctcgcacatggtcccaaggaagcattct 909
Db 202 GGGTTCAAGCTTTATAGACCCAGCATGATGTCTGACATGCGCCACAGGACCATTTCT 143
Qy 910 acattagtcaggagcctatctcttcaacaaagacatatacttccatcagcttaagg 969
Db 142 ACATTACTCAGGGGCTATCTGTATTAACAAAGACCATATATCTCTCATTTACCTTAAG 83
Qy 970 gcaacagcccttcaagcttctgtcagagcagcagtgagcagcgtgtgttgaat 1029
Db 82 GCACACAGCCCTTATATATATATATATATATATATATATATATATATATATATATAT 23
Qy 1030 aaacatgactgtatgaagaat 1051
Db 22 AAACATGACTGTATGAGAAAT 1

```

```

RESULT 5
LOCUS AW383386 477 bp mRNA EST 04-FEB-2000
DEFINITION PM3-HT0347-091199-001-B10 HT0347 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW383386
VERSION AW383386.1 GI:6888045
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 477)
AUTHORS HCGP http://www.ludwig.org.br/ORESTES.
TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

```

Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPSP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?pl1=PM3&ct2=PM3-PM347-091199-001-b10&ct3=1999-11-09&ct4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 338.  
Location/Qualifiers

## FEATURES

source

1. 477  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HT0347"  
/dev\_stage="Adult"  
/note="Organ: head, neck; Vector: puc18; Site: 1; Smal;  
Site: 2; Smal; A mini-library was made by cloning products  
derived from ORESSES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
BASE COUNT 122 a 115 c 102 g 138 t  
ORIGIN

Query Match 28.6%; Score 401.8; DB 114; Length 477;  
Best Local Similarity 92.9%; Pred. No. 6e-108;  
Matches 443; Conservative 0; Mismatches 32; Indels 2; Gaps 2;

517 catctgtcctaataaccagtaattgtctttaaagaagacatactatgcatc 576  
Db 477 CATTAATGCTCAAAAGTCCACCTATTCCTTTGAAGAAAGATATATGTCATT 418  
577 gggccagaacatccttgagacagctcccgagactgtctactgaccacagaaga 636  
578 gggccagaacatccttgagacagctcccgagactgtctactgaccacagaaga 636  
Db 417 GGGCCAGAACTTCAATGAGCAGCACTTACAGAGACCTGGTCACTGACCTCAGCAAGAA 358  
637 ggggtctcttccacaacaaagctcgaacgaacgaataatgcccagaagtggtc 696  
638 ggggtctcttccacaacaaagctcgaacgaacgaataatgcccagaagtggtc 696  
Db 357 GTGGGCTCTTCTGACACGAAAGCTGTCAAGCCAAATATATGACCAAGAGTGTT 298  
697 aggttgatctgcaagaaggaggtctccttgacaacattaccatctccacagccac 756  
698 aggttgatctgcaagaaggaggtctccttgacaacattaccatctccacagccac 756  
Db 297 AGGTTGATTCGAAAGGTAAAGGATTCCTCGACGACATTACCTCTACACAGCCAC 238  
757 atggtctgc-cctcttcgctgcagtgactggtgtgaggaaccaggaatgaagcg 815  
758 atggtctgc-cctcttcgctgcagtgactggtgtgaggaaccaggaatgaagcg 815  
Db 237 ATGGCTGCAATTTTTCGCTGCTAGTCATGTGCTACTAAGAACGAGATGAGAAAGTGG 178  
816 ctggccgatatatgtagcccgtaagtagaggaaggtctcaagctttagagccaggtg 875  
817 ctggccgatatatgtagcccgtaagtagaggaaggtctcaagctttagagccaggtg 875  
Db 177 CTGGCCAAATATATGAGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 118  
876 gtactccgcatgagcccaaggaagccattctacatagtcagggccatctcttaac 935  
877 gtactccgcatgagcccaaggaagccattctacatagtcagggccatctcttaac 935  
Db 117 GTATTCTGCGCATGCGCCAAAGGCAAGCCATTCTACATTAGTAGAGGCGCTATCTGTAA 58  
936 aaagacacatatatctcaatcttaagccttaaggaacagcccttacaattctc 992  
937 aaagacacatatatctcaatcttaagccttaaggaacagcccttacaattctc 992  
Db 57 AAAAGACCATATATCTCTCAATTCAGC-TTAAAGGCAACAGCCCTTATTAAGTTCT 2

RESULT 6  
AA435342 426 bp mRNA EST 04-AUG-1997  
LOCUS AA435342  
DEFINITION v49g09.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone  
IMAGE:803968 5' similar to WP:B0285.5 CE00644 ; mRNA sequence.  
ACCESSION AA435342  
VERSION AA435342.1 GI:2140162  
KEYWORDS EST.  
SOURCE house mouse.

## ORGANISM

## REFERENCE

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 426)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Giesel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HM Mouse EST Project  
Unpublished (1996)  
Contact: Maria M/Mouse EST Project  
WashU-HM Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
NCI:480312

## TITLE

Unpublished (1996)  
Contact: Maria M/Mouse EST Project  
WashU-HM Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
NCI:480312

## JOURNAL

Unpublished (1996)  
Contact: Maria M/Mouse EST Project  
WashU-HM Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
NCI:480312

## COMMENT

Unpublished (1996)  
Contact: Maria M/Mouse EST Project  
WashU-HM Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
NCI:480312

## FEATURES

source

1. 426  
/organism="Mus musculus"  
/strain="B6D2 F1/J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:803968"  
/clone\_lib="Knowles Solter mouse 2 cell"  
/tissue\_type="embryo"  
/dev\_stage="2-cell"  
/lab\_host="DH10B"  
/note="Organ: embryo; Vector: pBluescribe (modified);  
Site: 1; Muir; Site: 2; SalI; Cloned unidirectionally from  
mRNA prepared from 13,500 2-cell stage embryos. Primer:  
SalI (dr): 5'-CGGTGACCGTCCAGCGGTTTCTTTTCTT-3'.  
CDNAS  
were cloned into the Muir/SalI sites of a modified  
pBluescribe vector using commercial linkers (NEB).  
Average insert size: 1.2 kb."  
BASE COUNT 126 a 97 c 107 g 96 t  
ORIGIN

## FEATURES

source

1. 426  
/organism="Mus musculus"  
/strain="B6D2 F1/J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:803968"  
/clone\_lib="Knowles Solter mouse 2 cell"  
/tissue\_type="embryo"  
/dev\_stage="2-cell"  
/lab\_host="DH10B"  
/note="Organ: embryo; Vector: pBluescribe (modified);  
Site: 1; Muir; Site: 2; SalI; Cloned unidirectionally from  
mRNA prepared from 13,500 2-cell stage embryos. Primer:  
SalI (dr): 5'-CGGTGACCGTCCAGCGGTTTCTTTTCTT-3'.  
CDNAS  
were cloned into the Muir/SalI sites of a modified  
pBluescribe vector using commercial linkers (NEB).  
Average insert size: 1.2 kb."  
BASE COUNT 126 a 97 c 107 g 96 t  
ORIGIN

Query Match 25.9%; Score 364.2; DB 7; Length 426;  
Best Local Similarity 91.1%; Pred. No. 8.2e-97;  
Matches 387; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

589 tcatgagacacagttaccgagactgttcaactgacctcaaggaagagtggtcttcc 648  
590 tcatgagacacagttaccgagactgttcaactgacctcaaggaagagtggtcttcc 648  
Db 2 TCATGAGACACAGTTACCGAGACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 61  
649 aacacaaagctgtcgaagcaacaaagaataatgcccagaaggtgttagttgagcg 708  
650 aacacaaagctgtcgaagcaacaaagaataatgcccagaaggtgttagttgagcg 708  
Db 62 AACACAAAGCTGTCAAGCCAAACAAATCATGCCCAAAAGGTGTGTTAGTTGATTC 121  
709 aaaggaaggtctctcttgaacacattaccatctctacacagccacatgtgctc 768  
710 aaaggaaggtctctcttgaacacattaccatctctacacagccacatgtgctc 768  
Db 122 AAAGGGAAGGATTCCTCGACACATTTACCATCTCAACACAGCCACATGCTCATTC 181  
769 ttgcgtcagtgactgtgtgtgaggaacaggaatgagaaagggcgctggcgagttg 828  
770 ttgcgtcagtgactgtgtgtgaggaacaggaatgagaaagggcgctggcgagttg 828  
Db 182 TTTGCTGCAGAGTGTGCTGCTAGTAGAGAACAGAGATGAGAAAGGTGCGGCAATTTG 241  
829 gtgacccgtaagttaggggaaggtcttcaagctttagagcaggggtgtgactccgca 888  
830 gtgacccgtaagttaggggaaggtcttcaagctttagagcaggggtgtgactccgca 888  
Db 242 GTGACCCGGAAGTTAAGGGAAGGTTAAATCTTTTGAACACAGATGCTGCTGCAAG 301  
889 gcccaaggaagacacattctacatagtcagggccatctcttaacaaagacatata 948  
890 gcccaaggaagacacattctacatagtcagggccatctcttaacaaagacatata 948  
Db 302 GCACAAAGGCAAGCATCTCTACCTTAGTCAGGCGCTATCTTCTAACGAAGAATATGTA 361

QY 949 ttcctcaattcaagcttaagggcaacagcccttaacagtttctgcaagacagatga 1008  
 Db 362 TTCCTCAGTTCAGCTTAAGGCAACAGCCCTCATACAGTTTCCTGACAGACATGGA 421  
 QY 1009 gtcaa 1013  
 Db 422 GTTAA 426

RESULT 7  
 AA794925 455 bp mRNA EST 09-FEB-1998  
 LOCUS vrr4g04.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone  
 DEFINITION IMAGE:1123926 5', mRNA sequence.  
 AA794925  
 VERSION AA794925.1 GI:2857880  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 455)  
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Gelsel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.

TITLE The WashU-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL: contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:613262.

FEATURES  
 source location/Qualifiers  
 1..455  
 /organism="Mus musculus"  
 /strain="B6D F1/J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1123926"  
 /clone\_lib="Knowles Solter mouse 2 cell"  
 /issue\_type="embryo"  
 /dev\_stage="2-cell"  
 /lab\_host="DH10B"  
 /note="Organ: embryo; Vector: pBluescribe (modified);  
 Site: 1: Mui; Site: 2: Salt; Cloned unidirectionally from  
 mRNA prepared from 13,500 2-cell stage embryos. Primer:  
 SalI(drf): 5'-CGGTGACGCGCGACCGCTTTTCTTTT-3'.  
 were cloned into the Mui/Salt sites of a modified  
 pBluescribe vector using commercial linkers (NEB).  
 Average insert size: 1.2 kb."

BASE COUNT 138 a 97 c 118 g 102 t  
 ORIGIN

Query Match 24.7%; Score 346.6; DB 12; Length 455;  
 Best Local Similarity 88.0%; Pred. No. 1.4e-91;  
 Matches 402; Conservative 0; Mismatches 49; Indels 6; Gaps 2;

QY 594 gagcagattaccgggacctgtctactgacctcaggaagagtggtcttccacac 653  
 Db 5 GAGTACAGTTACAGAGAGGCTGCTACGACTCAGGAAGAGGCGCTGTCTTACAC 64  
 QY 654 aaaagctgtcaagcacaagaataatgcccagaaggtgtgtatgtatgtcgaag 713  
 Db 65 AAAAGCTGTCAAGCACAAGAAATCATGCCCCAAAAGGTGTTAGTTGATTCAAAAG 124

QY 714 gaaggtctcttgacacacattaccattctaccacagccacatgtgctcttcgc 773  
 Db 125 GAAGGATTCCTTGACACACATTCATCTCAACACAGCCACATGCGATCTCTTGC 184  
 QY 774 tgcagtgactgtgtgtgagaccagagatgaaagcgctgcccattatgtatc 833  
 Db 185 TGCAGTGACTGTGTGTGAGACAGATGAGAACAGATGAGAAAGTGCTG6CCATATGTGTAC 244  
 QY 834 ccgtaagttaggggaaggtctcaagtccttaagcagaaggtgtactccgcattgccc 893  
 Db 245 CCGGAAGTTAGGGGAAGGTTAAATCTTAAACACAGATGTACTCTGCATGAGACA 304  
 QY 894 agggcagccattctacattagtcagggcctatctcttaacaaaagccatattcct 953  
 Db 305 AGGCAAGCCATCTACTACTGTAGTCAGGGCCATATCTTTACGAAAGACTATGTATTC-- 362  
 QY 954 caattagcttaagggcaacagcccttacaagttctgtcagagacagatgacaa 1013  
 Db 363 TCAGTCAGCTTAAAGG---CACAGCCATACAAAGTTCCGTCAGACGATGAGTTAA 418  
 QY 1014 ggcgtgttataatgaataacatgactgtgtatgaaga 1050  
 Db 419 ACCCGTTCATGATTAACATGATGATGATGAA 455

RESULT 8  
 LOCUS CNS01MY 866 bp DNA GSS 12-MAY-2000  
 DEFINITION Tetradon nigroviridis genome survey sequence T7 end of clone  
 200J23 of library G from Tetradon nigroviridis, genomic survey  
 sequence.  
 A1170467.1 GI:7808524  
 GSS: genome survey sequence.  
 SOURCE Tetradon nigroviridis.  
 ORGANISM Tetradon nigroviridis.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; teleostei; Euteleostei; Neoteleostei;  
 Holacanthopterygii; Acanthopterygii; Percomorpha;  
 Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.  
 1 (bases 1 to 866)  
 Roest-Crolius, H., Jallion, O., Dasilva, C., Fizmes, C., Fisher, C.,  
 Bouneau, L., Billault, A., Queller, F., Saurin, W., Bernot, A. and  
 Weissenbach, J.  
 Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetradon nigroviridis  
 unpublished  
 2 (bases 1 to 866)  
 Roest-Crolius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,  
 Bernot, A., Fizmes, C., Winkler, P., Brotter, P., Queller, F.,  
 Saurin, W. and Weissenbach, J.  
 Human gene number estimate provided by genome wide analysis using  
 Tetradon nigroviridis DNA sequence  
 unpublished  
 3 (bases 1 to 866)  
 Genoscope.  
 Direct Submission  
 Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
 This sequence is a single read and was generated as part of a large  
 genome. For more information, please take a look at  
 http://www.genoscope.cns.fr/tetradon.

FEATURES  
 source location/Qualifiers  
 1..866  
 /organism="Tetradon nigroviridis"  
 /db\_xref="taxon:99883"  
 /clone="200J23"  
 /clone\_lib="G-"  
 /note="Genoscope sequence ID: COAG200CE12LP1-end : T7"

BASE COUNT 164 a 280 c 264 g 158 t  
 ORIGIN

Query Match	23.68;	Score 331.6;	DB 230;	Length 866;
Best Local Similarity	65.28;	Pred. No. 4.9e-87;		
Matches 561; Conservative	0;	Mismatches 224;	Indels 75;	Gaps 2.

[illegible]

RESULT	9
AI467374	
LOCUS	358 bp mRNA
DEFINITION	EST 09-MAR-1999
ACCESSION	U449609.x1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
	IMAGE:803968 3. similar to TR:O18756 O18756 C5-GIDUORONTL EPIMERASE
	;; mRNA sequence.
	AI467374

VERSION	AI467374.1	GI:4320/11
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	

## REFERENCE AUTHORS

REFERENCE	TITLE
1 (Pages 1 to 358)	
Marrin, M., Hillier, L., Kuwabara, T., Martin, J., Beck, C., Wylie, T., Pearson, Underwood, K., Steptoe, M., Phelps, B., Allen, M., Bowers, Y., Ritten, R., Swaller, T., Gibbons, M., M., T., Harvey, N., Schurr, R., F., Koh, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.	
the WashU-NCI Mouse EST Project 1999	
Unpublished (1999)	
Contact: Marria M/WashU-NCI Mouse EST Project 1999	
University School of Medicine	
COMMENT	

JOURNAL  
COMMENT

Washington, on: ewatson@wustl.edu  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouse@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:480312  
This clone was previously sequenced on the 5' end only, this new  
data is from the 3' end  
and is from the same clone. similarly on wrong strand  
Seq primer: Primer name ambiguous  
High quality sequence stop: 352.

## FEATURES

```

/organism="Mus musculus"
/strain="B6D2 F1/J"
/db_xref="taxon:10090"
/clone_image="803968"
/clone_lib="Knockles Solter mouse 2 cell"
/tissue_type="embryo"
/dev_stage="2-cell"
/lab_host="DH10B"
/notes="Organ: embryo; Vector: pbluescribe (modified);
Site_1: Muir; Site_2: Salt; Cloned unidirectionally from
mRNA prepared from 13,500 2-cell stage embryos. Primer:
Salt(dT): 5'-CGCTCACCCTCGACCGTTTTTTTTTTT-3'.
CDNAS
were cloned into the Muir/Salt sites of a modified
pbluescribe vector using commercial linkers (NEB).
Average insert size: 1.2 kb."

```

Overall Match 21.7%; Score 304.2; DB 20; Length 358;

Overall Match	90.8%;	Pred. No. 4.8e-79;	
Best Local Similarity	90.8%;		
Conservative	0;	Mismatches 33;	Indels 0;
			Gaps 0;

Matches 324 / CONSERVATIVE 641

[illegible][illegible]

642 tcttccacacaaagctgtccaagccaacagaataatgcccagaagryllylccaglc

Db

61 CCTTCTAACACAAAGCTGTCAAGCCAACCAAAATCATGCCCAAAAGGTGGTTAGGTT 120

702 gattgcgaagggagggcttcctgcacaacattaccatctaccacagccacatgyc 761

Db 121 GATTGCCAAAGGGAAGGATTCTTGACAACATTACCATCTCAACCACAGCCACATGGC 180

762 tacccttcgcctaccgaatgaactgctggtgaggaaccagatgagaagcgctggcc 821

101 mcaatctttctctgaagtgactggcctagtggacacacgattgagaagcttgccctggcc 240

DB  
101 19CMTCTTCCGCGGGAGTTTCAGATCTTTAGACCAAGGTGTACTC 881

[illegible]

**D**b      241 AATTATGCTACCCGGAGTTAGGGGAAGCGTAAATCCTAAGAACAATA.....



QY 882 cgcacatgcccagggcgaagcatttctacatttagtcaaggcctatcttaacaa 938  
 Db 301 tcccatggcacaaggcgaagcatttctacatttagtcaaggcctatcttaacaa 357

RESULT 10  
 LOCUS BE170776 434 bp mRNA EST 1 21-JUN-2000  
 DEFINITION OVA-HT0539-160300-137-e07 HT0539 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BE170776  
 VERSION BE170776.1 GI:8633497  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 434)  
 Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Brlones,M.R.,  
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
 Brunsteln,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
 'M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 CONTACT: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=8t2-qv4-HT0539-160  
 300-137-e07&ts=2000-03-16&cl=1)  
 Seq primer: puc 18 forward  
 High quality sequence stop: 434.  
 Location/Qualifiers  
 1. 434  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="HT0539"  
 /dev\_stage="Adult"  
 /note="Organ: head,neck; Vector: puc18; Site:1: SmaI;  
 Site:2: SmaI; A mini-library was made by cloning products  
 derived from ORESSES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."  
 BASE COUNT 122 a 87 c 87 g 138 t  
 ORIGIN

Query Match 21.6%; Score 303; DB 163; Length 434;  
 Best Local Similarity 91.5%; Pred. No. 1.2e-78;  
 Matches 321; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 360 tgttaacagcttattgctccaggaacgtaagtgatctcttgcgaactgggaagac 419  
 Db 363 tctttgattatcttcccttaccagacgtaagtgatcttgcgaactgggaagac 304

QY 420 aaaaattatatttatttatttgcctcaagttcttaacaaatgaaagcgtgtgtgt 479  
 Db 303 AAAAGATTATATATTTGACTTGCACCTCAAGTTCTTGAACAATGSAATGTCCTGCT 244

QY 480 tctggagagcaacaaatcaagcttctcctgtacattatgtcnaatcaccagct 539  
 Db 243 TCTAGAGACCAAGAAATACCTCTTCTACATACATTATGTCNAATGCTCAGACT 184

QY 540 aattcctttaagaagaagacatatattatgcatcgtggcccaagaacatcagagcac 599  
 Db 183 AATTGCTTTTAAAGAAAGATATATACTATAGTCATTGGGCCAGAACTTCATGAGAC 124

QY 600 agttaccggagcctgtcactgaccccaagaagaagtggtgtcttccaaacaaagc 659  
 Db 123 AGTTACCAAGGAGACCTGTGCTACCTGACCTCAGAAAGAGTGCTTTTCAACACAAAGC 64

QY 660 tgtcaagcccaagaatgatcccaagaagtggtgtatgtatgtcgaa 710  
 Db 63 TGTCAAGCCAAACCAAAATATGATGCCAAGAGTGCTAGGTGATTCGAAA 13

RESULT 11  
 LOCUS BF080558 566 bp mRNA EST 18-OCT-2000  
 DEFINITION 231598 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.  
 ACCESSION BF080558  
 VERSION BF080558.1 GI:10874460  
 KEYWORDS EST.  
 SOURCE pig.  
 ORGANISM Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 REFERENCE 1 (bases 1 to 566)  
 Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,  
 Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.  
 and Keele,J.W.  
 Design and use of two pooled tissue normalized cDNA libraries for  
 EST discovery in swine  
 Unpublished (2000)  
 CONTACT: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4390  
 Fax: 402 762 4390  
 Email: smith@email.marc.usda.gov  
 Single pass sequencing. Bases called and alt\_trimmed with phred  
 v0.980904.e. Vector identified by cross\_match with the -m1nscore 18  
 and -mismatch 12 options.  
 PCR Primers  
 FORWARD: AGCAAAACAGCTATGACCAAT  
 BACKWARD: GTTTCACGATCAGCAGC  
 Plate: 40 row: B column: 6  
 Seq primer: ATTATGCGTACACTATAG.  
 Location/Qualifiers  
 1. 566  
 /organism="Sus scrofa"  
 /db\_xref="taxon:9823"  
 /clone\_lib="MARC 2P1G"  
 /tissue\_type="pooled"  
 /lab\_host="PH10B"  
 /note="Vector: PCMV SPOR6; Site:1: XbaI; Site:2: XhoI;  
 Library made from pooled tissue from testis, ovary,  
 endometrium, hypothalamus, pituitary, and placenta."  
 BASE COUNT 182 a 107 c 135 g 142 t  
 ORIGIN

Query Match 20.3%; Score 285.4; DB 143; Length 566;  
 Best Local Similarity 94.9%; Pred. No. 2.1e-73;  
 Matches 295; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 13 tctcaatgatttcaaaagtctatcagagagagcccttataacctgagtgtgttt 72  
 Db 256 TCTCATAGCTATTCCAAAGCTCTATACAGAGAGCCCTTATACCCCTGATGCTGTT 315

QY 73 atgtccttgaagctacaaatgtggaagtcgagacagagcgaagtgcataaagtgggtt 132  
 Db 316 ATGTCTTGAAGGCTACAAATGTGAGAGTCGAGACAGCAAGTCAATGAGGGGTT 375

QY 133 gaagtgatccttattatcacacagtgaggacccaaggtatttctacccaatccagatt 192



```

|||||
Db 376 GAAGGTGACCTTGTCTACACATGGGACCTCAAGGCTATTCTTACCACATCCAGATT 435
Qy 193 gacagatagggttaagttactacacagcaagaatcctaactgaaaaaccccccatataag 252
Db 436 GCACACTATGAGATTAACTATTACATAGAAATCTACTGCAAAACCCCTCACATAGAA 495
Qy 253 gtaataaagacagcaagaagagcaaaaacagcaagcccaatgactgactgagcc 312
Db 496 GTATATGAACACAGCAGACAGGAGCAAAATACAAAGCTAGTACTGACTGTGCA 555
Qy 313 aagagctgctt 323
Db 556 AAGGCTGCTT 566

RESULT 12
LOCUS A1981391 590 bp mRNA EST 10-JUN-2000
DEFINITION pat.pk0052.c6.f chicken activated T cell cDNA Gallus gallus CDNA
clone pat.pk0052.c6.f 5' similar to KIAA0836 protein. mRNA
sequence.
ACCESSION A1981391 GI:5884419
VERSION A1981391
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 590)
Tirunaguru, V.G., Sofer, L., Cul, J. and Burnside, J.
An expressed sequence tag database of T-cell-enriched activated
chicken splenocytes: sequence analysis of 5251 clones
JOURNAL Genomics 66 (2), 144-151 (2000)
MEDLINE 20318616
COMMENT Contact: Joan Burnside
Molecular Endocrinology
University of Delaware
40 Townsend Hall, Newark, DE 19717, USA
Tel: 302 831-1345
Fax: 302 831-3651
Email: joan@udel.edu
Seq primer: T7.

FEATURES
Source location/Qualifiers
1..590
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="pat.pk0052.c6.f"
/clone_lib="chicken activated T cell cDNA"
/sex="male"
/cell_type="con A-activated splenic T cell"
/lab_host="E.coli TOP10 F"
/note="Vector: pCDNA3"

BASE COUNT 146 a 165 c 160 g 113 t 6 others
ORIGIN
Query Match 20.38; Score 285.4; DB 104; Length 590;
Best Local Similarity 76.08; Pred. No. 2.2e-73;
Matches 352; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy 119 gcaataagtgagggtgaaggtgtacctatctacacagtgaggacctcaagctattctc 178
|||||
Db 1 GCATATAGTGTGTTGTAAGGTGTGCTTATTCACACCCAGTGGGACCTCCAGGGTACTCT 60
Qy 179 acccaatccagatgacagatgagttgaagttacacacagcaagaatcctaactgaaaac 238
|||||
Db 61 ATCCCATCCAGATGCACAGTACGAGTTGAGTCACTACAGCAAAACCTGACGAGAGAGC 120
Qy 239 cccctcatatgaggttatataaagcagcagaagcagagcaaaaacaggaagcccaatg 298
|||||
Db 121 CCCCACATCGAGGTGTGTAAGCAAGGCCGAGGAGGAAGGAGGAGGAGGAGGAGGAGGAGG 180

```

```

Qy 299 actgagctgtgcccaaggcgtgctttagtgctagtgtgctgataagtaagtaagtaacca 358
|||||
Db 181 AGTGACGGGTGCCCCAAAGGCTGCTCCCTTCCACAGGTGCCGATTAAGCCAAATTACCA 240
Qy 359 atgttaaacagttcctgctcagaaaccagtgaaagtgatccttgaactggggagaca 418
|||||
Db 241 GTCTTAACACTTTGTGTGCTCCAGAACACCGAGGGGTTTCCCTGAGCTTGGCAAGC 300
Qy 419 caaagatttattatcttcaattgacctcaagttcttaacaatggaaggtgctgtg 478
|||||
Db 301 CCAGAGATTTCATTATTCTTTCGATCTCAAGCTCTGTAACCAAGCGGAGCAATTTCTGTGG 360
Qy 479 ttctgagacagcagaaagaatcagctcttcactgatactatgctcaaatccagc 538
|||||
Db 361 TGCTGAGACCCAGGAGAAACCAAGCTTTTACCGCTGACTAGCTCTCCAAACGAGCAGC 420
Qy 539 taatgctttaaagaagaacataactactatgcatcgagcc 581
|||||
Db 421 TCATCGCCTTACGAGGAGCGCAGCATCTACTACGATCGGAGGCC 463

RESULT 13
LOCUS BE861166 381 bp mRNA EST 29-SEP-2000
DEFINITION UI-M-AM0-adu-h-07-0-UI.r1 NIH-BMAP_MAM Mus musculus CDNA clone
UI-M-AM0-adu-h-07-0-UI 5', mRNA sequence.
ACCESSION BE861166 GI:10378842
VERSION BE861166
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 381)
Bonaldi, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chih, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: MEST@mail.nih.gov
cDNA library Preparation: M.B. Soares Lab Clone distribution:
Researchers may obtain BMAP CDNA clones from RESEARCH GENERICS. It
should be noted that Bento Soares is generating a small number of
additional specialized non-redundant arrays of BMAP cDNAs whose
availability will be considered under appropriate and limited
collaborative arrangements
Seq primer: M13 Reverse.

FEATURES
Source location/Qualifiers
1..381
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-AM0-adu-h-07-0-UI"
/clone_lib="NIH-BMAP_MAM"
/dev_stage="7-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NIH-BMAP_MAM library is a non-normalized library
constructed from mouse amygdala. The tag is a string of 5
nucleotides present between the Not I site and the
oligo-dT track. The library was constructed as described
by Bonaldi, Lennon and Soares, Genome Research 6: 791-806
, 1996. Tissue provided by Ms. Anne Novakovich,
Zivic-Miller Laboratories."

BASE COUNT 100 a 110 c 82 g 89 t

```



## COMMENT

Contact: Nori Satoh  
Department of Zoology  
Kyoto University  
Sakyo-ku, Kyoto 606-8502, Japan  
Tel: 81-75-753-4081  
Fax: 81-75-705-1113  
Email: satoh@esclidian.zool.kyoto-u.ac.jp.

## FEATURES

## source

location/Qualifiers  
1. 684  
/organism="Clona intestinalis"  
/db\_xref="taxon:7719"  
/clone="clb2b15"  
/clone\_id="Nori Satoh unpublished cDNA library"  
/tissue\_type="whole animal"  
/dev\_stage="tailbud"

BASE COUNT 230 a 113 c 145 g 196 t  
ORIGIN

Query Match 10.2%; Score 143.6; DB 31; Length 684;

Best Local Similarity 52.1%; Pred. No. 2.6e-31;

Matches 345; Conservative 0; Mismatches 314; Indels 3; Gaps 1;

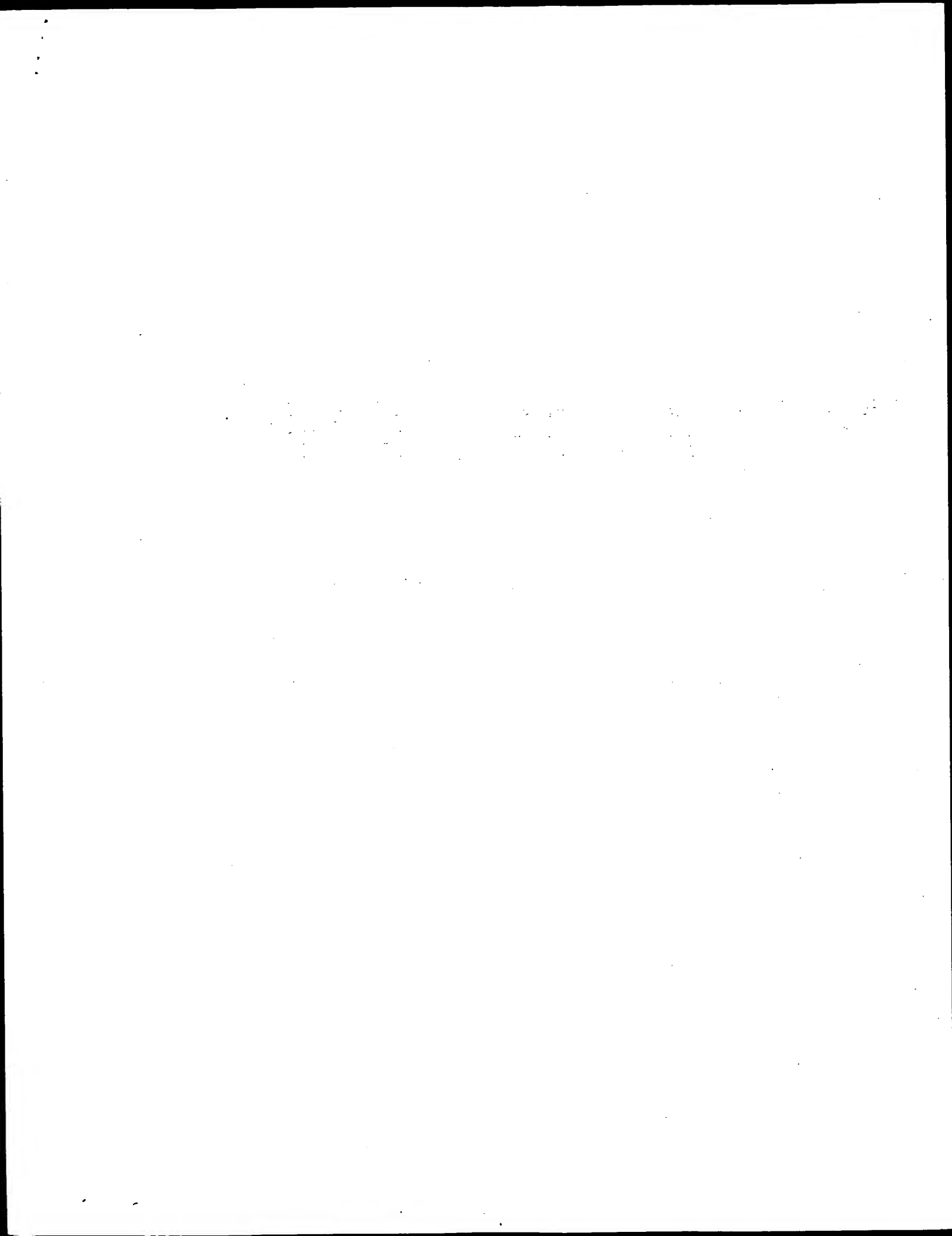
```
Oy 165 tcaaggctatctctaccatccagatgacagatggttaagtcctacacaaaga 224
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4 tgaaggttactatgtatggaatccagattgctcagttggcattgactacacaaaca 63
Oy 225 tctaactgaaaaaccccccatagaggtatgaaacagagaagacaagaa 284
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 64 catgacacaaagccacacaaacaaacagtttatgaaatgacaggggtctgctcagtc 123
Oy 285 cagcaagcccaatgactgactgagcccaaggctgcttattgctagtgctgataa 344
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 124 tcggtgctctccagaaatgaaatgctgttggtgaaacgaagaagatgaaagaa 183
Oy 345 gtcaagattaccacatggttaaacagttcctcagaacacagtgaggtgtacctt 404
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 184 ttcttcgagttatgaaatgacactccagattctattccattgggaagtggtccct 243
Oy 405 gcaactgggaacacaaagatttatttcaattgacctcaagttcctaacaatg 464
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 244 tgttcttggaacacacacagtttatgctggttgatgattgaaatggtgcaaatgg 303
Oy 465 aagcgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 524
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 304 ttcttattctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 363
Oy 525 ctcaataccacgactaattgcttta---agaagaacatactatgcatcgagcc 581
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 364 cacaacacagttggtctatttattgatttgaatcaaatatatttaccgaatcggtcc 423
Oy 582 cagacacatcagcagcagttaccgagacgtgctgctgctgctgctgctgctgct 641
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 424 aagtcgattgacatgacatgacagaggtggtgacagatttgaaagaggtattgg 483
Oy 642 tcttccaacacaaagctgctcaagccaaagaataatgcccagaagtggttaggt 701
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 484 tcttacaactgaaacaaacaaagaaagttgaaattgcccattcannaaggtttcaagcct 543
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 702 gattcgagaagggagggctctctgacaacattacatctacacagccacacagc 761
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 544 tatgttaaggggtggtggttatattgattatgacacgtccacgctgacacattgac 603
Oy 762 tgccttctcgtcgcagtgactgctgctgctgctgctgctgctgctgctgctgct 821
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 604 ccaattttttgatgacacacacacacacacacacacacacacacacacacacacac 821
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 822 ga 823
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 664 ga 665
```

Search completed: May 25, 2001, 03:23:20  
Job time: 5137 sec

Fri May 25 10:28:27 2001

us-09-403-269-12\_copy\_1\_1404.rst

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2001, 02:20:59 ; Search time 1798.43 Seconds

(without alignments)  
11513.372 Million cell updates/sec

Title: US-09-403-269-12\_COPY\_1\_1404

Perfect score: 1404

Sequence: 1 tccaagctgaattctcatag.....gcagcgggcaagcacac 1404

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_ba3: \*  
4: gb\_in1: \*  
5: gb\_in2: \*  
6: gb\_in3: \*  
7: gb\_om: \*  
8: gb\_ov: \*  
9: gb\_pat1: \*  
10: gb\_pat2: \*  
11: gb\_ph: \*  
12: gb\_p11: \*  
13: gb\_p12: \*  
14: gb\_p13: \*  
15: gb\_p14: \*  
16: em\_ba1: \*  
17: em\_ba2: \*  
18: em\_fun: \*  
19: em\_htgo\_hum: \*  
20: em\_htgo\_inv: \*  
21: em\_htgo\_rnd: \*  
22: em\_htg\_hum1: \*  
23: em\_htg\_hum2: \*  
24: em\_htg\_hum3: \*  
25: em\_htg\_hum4: \*  
26: em\_htg\_hum5: \*  
27: em\_htg\_hum6: \*  
28: em\_htg\_hum7: \*  
29: em\_htg\_hum8: \*  
30: em\_htg\_inv1: \*  
31: em\_htg\_inv2: \*  
32: em\_htg\_other: \*  
33: em\_htg\_rnd: \*  
34: em\_hum1: \*  
35: em\_hum2: \*  
36: em\_hum3: \*  
37: em\_hum4: \*  
38: em\_hum5: \*  
39: em\_hum6: \*  
40: em\_hum7: \*  
41: em\_in: \*  
42: em\_om: \*  
43: em\_or: \*

44: em\_ov: \*  
45: em\_pat: \*  
46: em\_ph: \*  
47: em\_pl: \*  
48: em\_ro: \*  
49: em\_sts: \*  
50: em\_sy: \*  
51: em\_un: \*  
52: em\_v1: \*  
53: gb\_sts1: \*  
54: gb\_sts2: \*  
55: gb\_sts3: \*  
56: gb\_sy: \*  
57: gb\_un: \*  
58: gb\_v11: \*  
59: gb\_v12: \*  
60: gb\_htg1: \*  
61: gb\_htg2: \*  
62: gb\_htg3: \*  
63: gb\_htg4: \*  
64: gb\_htg5: \*  
65: gb\_htg6: \*  
66: gb\_htg7: \*  
67: gb\_htg8: \*  
68: gb\_htg9: \*  
69: gb\_htg10: \*  
70: gb\_htg11: \*  
71: gb\_htg12: \*  
72: gb\_htg13: \*  
73: gb\_htg14: \*  
74: gb\_htg15: \*  
75: gb\_htg16: \*  
76: gb\_htg17: \*  
77: gb\_htg18: \*  
78: gb\_htg19: \*  
79: gb\_htg20: \*  
80: gb\_htg21: \*  
81: gb\_htg22: \*  
82: gb\_htg23: \*  
83: gb\_htg24: \*  
84: gb\_htg25: \*  
85: gb\_pr1: \*  
86: gb\_pr2: \*  
87: gb\_pr3: \*  
88: gb\_pr4: \*  
89: gb\_pr5: \*  
90: gb\_pr6: \*  
91: gb\_pr7: \*  
92: gb\_pr8: \*  
93: gb\_pr9: \*  
94: gb\_r01: \*  
95: gb\_r02: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1404	100.0	3085	7 AF003927	AF003927 Bos tauru
2	1249.6	89.0	4721	89 AK002188	AK002188 Homo sapi
3	1249.6	89.0	4791	85 AB020643	AB020643 Homo sapi
4	1203.2	85.7	2088	94 AF325532	AF325532 Mus muscu
5	933	66.5	167187	71 AC026992	AC026992 Homo sapi
6	490.6	34.9	171073	75 AC073380	AC073380 Homo sapi
7	204	14.5	167187	71 AC026992	AC026992 Homo sapi
8	150.8	10.7	29372	65 AC018117	AC018117 Drosophi
9	150.8	10.7	119597	60 AC007593	AC007593 Drosophi
10	150.8	10.7	167062	60 AC007624	AC007624 Drosophi
11	150.8	10.7	299474	4 AE003789	AE003789 Drosophi





```

Db      841 TTAGGGGAGGCTTCACTTTTAGAGCCAGGAGTACTCCGCCATGCGCCCAAGGCGAA 900
Qy      901 gccatttacaatagtaagggcctatctcttaacaaagaacatattcttaattca 960
Db      901 GCCATTCTACATTAGTAGAGGCTCTATCTTTAAACAAAAGCCATATATCTCTTAATTC 960
Qy      961 gcttaagaggaacagacccttacaagttctctgacagagcagatggagtcagtcgtg 1020
Db      961 GCTTTAAGGCGAACAGCCCTTACAGATTCTGTACAGAGCATGGAGTCAAGCTGTG 1020
Qy      1021 ttatgataaacaatagctgtagaagaataatccaactaactaactaactctttgttta 1080
Db      1021 TTTATGATTAACATGATGATGATGAAGAAATATCAATCACTACCTCTTTTGTGTTTA 1080
Qy      1081 aatgcttattatattcttattatgagctgtagtaacttaaaagaactgcaaggggaaaaa 1140
Db      1081 AATGCTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140
Qy      1141 ctgaggaagaagagagagtcctctgtatgagcgtgagcagatggaatcccttaagccatgctc 1200
Db      1141 CTCGGGAAAGAGAGAGAGCTCTGTGTATGAGCTGGCATGGAATCCCTTAAGCCATGCTC 1200
Qy      1201 cccctgtagcagactgctgtagaagacatctatgacctccgagcattcattgctgatt 1260
Db      1201 CCCTGTATGACGACTGCTGCTCAAGAACATATATGACCTCCGGCATCTTCATGCTTGGCATT 1260
Qy      1261 gccccaactgagcagctgtagaactatcacacacacacacacacacacacacacacacacac 1320
Db      1261 GCCCCCAACTGAGCCCTGGGACTATCACACACACACACACACACACACACACACACACAC 1320
Qy      1321 agcacctatgtagtcccaatcttaagaattgtcaagaggtggaagagctactc 1380
Db      1321 AGCACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1380
Qy      1381 aaagcagcgcgggcaagaacacacac 1404
Db      1381 AAAGCAGCGCGGCAAGACACAC 1404

RESULT 2
LOCUS   AK002188      2121 bp      mRNA      PRI      22-FEB-2000
DEFINITION Homo sapiens cDNA FL11326 fis, clone PLACE1010481, highly similar
to Homo sapiens mRNA for KIAA0836 protein.
ACCESSION AK002188.1 GI:7023912
VERSION   1.0
KEYWORDS   capping; fls (full insert sequence).
SOURCE     Homo sapiens placenta cDNA to mRNA, clone_1lb:PLACE1
           clone:PLACE1010481.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (sites)
AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
           Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
           Wadatsuma,M., Hosokiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
           Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
           Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
           Ishii,S., Kawai,Y., Saito,K., Yamamoto,D., Wakamatsu,A.,
           Nakamura,Y., Nagahara,K., Masuno,Y., Nimomiya,K. and Iwayanagi,T.
           NEDO human cDNA sequencing project
           Unpublished (2000)
           2 (bases 1 to 2121)
TITLE     Isogai,T. and Otsuki,T.
REFERENCE 16-FEB-2000) to the DBJ/EMBL/GenBank databases. Takao
           Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana,
           Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
           Tel:81-438-52-3951, Fax:81-438-52-3952)
           NEDO human cDNA sequencing project supported by Ministry of
           International Trade and Industry of Japan; cDNA full insert
           sequencing: Research Association for Biotechnology; cDNA library

```

```

construction, 5' - 3' end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
    source
        1..2121
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="PLACE1010481"
        /tissue_type="Placenta"
        /note="cloning vector: pHE18FL3"
BASE COUNT  654 a 411 c 453 g 603 t
ORIGIN
Query Match      89.0%; Score 1249.6; DB 89; Length 2121;
Best Local Similarity 94.1%; Pred. No. 0;
Matches 1310; Conservative 0; Mismatches 79; Indels 3; Gaps 1;
Qy      13 tctcattgattccaaagtctatgacagagagcccttaccctgtagtggtt 72
Db      52 TCTCATAGCTATTTCCTCAAGTCTATGACAGAGAGCCCTTACCCGATGCTGTTT 111
Qy      73 atgtccttgaagagctacatgttgaaagtcgagagacagagtagatgtagtggtt 132
Db      112 ATGTCTTTGAAGGCTACATGTGGAAGTCCAGACAGAGTGAATGATAGTGGGTT 171
Qy      133 gaagtgtagccttattctacacagctgggagcctcaaggtattcttaaccaatcagaat 192
Db      172 GAAGGTGTCCTATTCATACCAATGGGGAGCTCAAGCTATTCTATCCATCCAGATT 231
Qy      193 gacagatagggttagtactacagagagaaatctactgaaacacccctcatatagag 252
Db      232 GCACAGTATGATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAG 291
Qy      253 gtatatgaacacagcagagagacagagacaaacacagcagcccaatgtagctgctc 312
Db      292 GTATATGAAGACAGCAGAGAGACAGACAAACAAACAAACAAACAAACAAACAAAC 348
Qy      313 aagagctgctttagtgcagtagtggtgtagatgaatgaatgaatgaatgaatgaatgaatgaat 372
Db      349 AAGGGCTGCTTTATGCGGATGTGCTGATTAAGTCTATTAAGTCTATTAAGTCTATTAAG 408
Qy      373 attgtccagaaacaggagagagtgatcctctgcaacgctgggagacacaaagattatt 432
Db      409 ATTGCACAGAAACAGCAAGTATGCTCTTGCACACTGGGAACACAAAGATTATTAT 468
Qy      433 attcatttgacctcaagtlcttaacaaatggaagcgtgctgtgtctgtgtagagcaga 492
Db      469 ATTTCAATTTGACCTCAAGTTCTTGCACAAATGGAAGTGTGCTGGTGTCTAGAGCCACA 528
Qy      493 gaaagaatcagctcttctacgtagatattatgtctcaaataccagcgaattgtcttaaa 552
Db      529 GAAAGAAATACGCTCTTCACTATACATTAATATGTCCTCAAGTCTCAAGTCTTTTAA 588
Qy      553 gaaagagacataactatgacatcgagccagaaacatcatatgagacagagtagccggagc 612
Db      589 GAAAGAGATATATACTATAGGATGGGCCAGACACTTATGAGACACAGTTACAGGAGAC 648
Qy      613 ctgttactgacctcaggaagagtggtgtcttccacacacaaagctgtcaagccaca 672
Db      649 CTGTCACTGACCTCAGCAAGAGAGTGGTCTTCAACACAAAGCTGTCAAGCCAAC 708
Qy      673 agaataatgcccagaagagtggttaggttgtagtcgaagaagagagcttccctgaacac 732
Db      709 AAAATATATGCCCAAGAGAGGTGATGATGATGATGATGATGATGATGATGATGATGATG 768
Qy      733 attaccatctacacacagccacatgctgctctctgctgacagtagctgagtggtt 792
Db      769 ATTACCATCTCTACACAGCCACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 828
Qy      793 aggaaccaagatgagaaagcgcgtgacgattatgtagcccgtagtaagtagggaagc 852

```

FEATURES	LOCATION/Qualifiers
Db	AGGAGCAGGATGAGAAAGTGGCTGGCCAAATTATGTTGACCCGTAGTTAAGGGAGCG 888
Qy	ltcaagtccttaagagccagggttgactccgcgccaatggcccaaggaggaagccattctaca 912
Db	TTCAAGCTTTTAAAGGCCAGAGTGGTATTTCGCCATGGCCCAAGGGCAAGGCATTTTCACA 948
Qy	tttagtaaggccctatctccttaacaaagacatatatctcccaattcaagctttaagagca 972
Db	TTAATCAGGGCCCTATCTGTTTACAAAAGACCATATATCTCTCAATTCAGCTTTAAGGGCA 1008
Qy	acagccccccttaacaaagcttcgtctcagagcagcaatggagltcaagagctgtgtatgaataa 1032
Db	ACAGCCCCCTTTAAGTTCATCTTCATCGACAGCATGGAGTAAAGCTGCTTTATATAATAAAA 1068
Qy	catgacccggtatagaagaattccaaactaacccctagccctttgttttaaatggcttatg 1092
Db	CATGACTGGTATGAAAGATATCCAAACACACACTAGCTCTTTGTTTAAATGGCTTTATG 1128
Qy	tatcttcaatctggcgtgtatgacttaaaagaacatcgaggagaaacatcggagaa 1152
Db	TATCTTTAATTTGGGCGTGTATGACTTAAAGAAGACTGACGGGAAAAAATCGGAAAAAGAA 1188
Qy	gcgaggtctctgtatgagcgtgagcatggaaatcccttaagaagcatgctcccttttagac 1212
Db	GCAAGGCTCTTGTATGACCGGTGCATGGAAATCTCTTAAAGCATAGCTGCCCCGTGATGAC 1248
Qy	actggtcagaagaacatctatgacctccgagcatctcatgctctggcatgtgcccccaactg 1272
Db	ACTGGCTTAGGAACCATCTATGACCTCGCTCATCTTCATGCTTTGGGATGGCTCTTAACCTG 1308
Qy	gcccgtctggagctatcacaccacccaatcaataactgcagcgtcttaagacattgat 1332
Db	GCTCGCTGGGACTATGATCAATACACACCACTCATCATGATGTCAGCTACATGATCAATTTGAT 1368
Qy	gagtcgcccaacttcaagaagattgtcaagggtgtgaagaagctacacttaagaagcagccgg 1392
Db	1369 CAGTGCCCAATCTTCAAGAATTTGTCTCAAGAGGTGGAAAGAACTTAAGGGCAGCAGG 1428
Qy	1393 gcaaaagcacaaac 1404
Db	1429 GCAAAAGCACAAAC 1440
RESULT 3	
AB020643	LOCUS AB020643 4791 bp mRNA PRI 16-JUN-1999
DEFINITION	Homo sapiens mRNA for KIAA0836 protein, partial cds.
ACCESSION	AB020643
VERSION	AB020643.1 GI:4240160
KEYWORDS	
SOURCE	Homo sapiens adult male brain cDNA to mRNA, clone_1lb:pb1uescript1111
ORGANISM	SK plus clone:hj06165.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE	1 (sites)
JOURNAL	Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Hirosewa,M., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
MEDLINE	Prediction of the coding sequences of unidentified human genes.
AUTHORS	XII. The complete sequences of 100 new cDNA clones from brain which
REFERENCE	code for large proteins in vitro
JOURNAL	DNA Res. 5 (6), 355-364 (1998)
ABSTRACT	99156230
KEYWORDS	2 (bases 1 to 4791)
REFERENCE	Ohara,O., Suyama,M., Kikuno,R., Nagase,T. and Ishikawa,K.
AUTHORS	Direct Submission
JOURNAL	Submitted (02-DEC-1998) to the DBJ/EMBL/GenBank databases. Osamu
ABSTRACT	Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology, Japan
KEYWORDS	Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
REFERENCE	(E-mail:cdna@fookazusa.or.jp, Tel: +81-438-52-3913,
ABSTRACT	Fax: +81-438-52-3914)
KEYWORDS	Location/Qualifiers

[illegible]

Db 977 GAAAGATATATATATGATGCGATTTGGCCGACAACTTCATGACGACACATGACAGGAC 1036  
 Qy 613 ctgtctactgacctcgaagaagagtggtctcttcccaacaaagctgtcaagccaaca 672  
 Db 1037 CTGTCTACTGACCTCAGGAAGAGTGGCTCTTCAACAAACAAAGCTGACACCAACC 1096  
 Qy 673 agataatgcccagaagagtggttaagttgatttcggaagaagagcttccctgacaac 732  
 Db 1097 AAAAAATATGCCCCAGAAAGAGTGGTGAATGCAAAAGGATTCAGGATTCCTGACCAAC 1156  
 Qy 733 attacactctacacacagccacatgctgctctctctctctctctctctctctctctg 792  
 Db 1157 ATTAGCATCTCTACACAGCCACATGGCTGCTCATTTTCTGCTGCTGCTGCTGCTGCT 1216  
 Qy 793 aggaacacagatgagaagagcgctgacgattatgtgacccgtaagtaggaggaagc 852  
 Db 1217 AGGACACAGATGAGAAAGGCGCTGGCCAAATATGCTGACCCCTAGAGTGAAGGAGG 1276  
 Qy 853 ttcaagctcttaagagcagagtggttactccgacatgcccagagggagccattctaca 912  
 Db 1277 TTCAAGCTTTAGAGCCAGAGTGGTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1336  
 Qy 913 tttagcagagccttactctcttcaacaaagacatattctctcaattcagctttaagggca 972  
 Db 1337 TTACTCAGGGGCTATCTGTTAACAAGAACCATATATCTCTCAATGACCTTTAAGGCA 1396  
 Qy 973 acagcccttacaagattctgtcgaagcagatggagtaagagctgtgttataatgaa 1032  
 Db 1397 ACACCCCTTATAGTTTCTATGACACAGCATGAGAGTAAAGCTGCTGCTGCTGCTGCT 1456  
 Qy 1033 catbacggtatgaagaatctcaactacacactacactctctctctctctctctctct 1092  
 Db 1457 CATGACTGATGAGAAATATCCACACACACTGCTCTTTGTTTAAATGAGCTTTATAG 1516  
 Qy 1093 tattcttaatttggtgtgtatgacttaaaagaaactgcaaggggaaacacgcggaagaa 1152  
 Db 1517 TATCTTAAATGAGGCTGATGACTTAAAGAACTGACAGGGGAAAAACCTCGGAAAAAG 1576  
 Qy 1153 gcaagagctctgtatgagcgtgacggaatcccttaagacatgctccctctgtacgac 1212  
 Db 1577 GCAAGGCTCTGTATGAGCGCTGACATGGAATCTTAAACCCATGCTGCTGCTGCTGCT 1636  
 Qy 1213 actggtcagaagacatctatgacctccgacactcatgctgtgacgtgcccccaacg 1272  
 Db 1637 ACAGGCTGAGGAACATCTATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1696  
 Qy 1273 gcccgctggactatcaac 1332  
 Db 1697 GCTCGCTGGGACTATGATACACACACACATCAATGCTGAGCTACTGATCAATGAT 1756  
 Qy 1333 gactcccaactcttcaagaatttgcagaaggtggaagagctacctaagaagcagcg 1392  
 Db 1757 GAGTCCCAATCTTCAAGAAATTTGCAAGAGGTGAAAAAGCTTAAGGCAAGCAGG 1816  
 Qy 1393 gcaaaagcacaac 1404  
 Db 1817 GCAAGACACAC 1828

RESULT 4  
 AF325532 2088 bp mRNA ROD 22-DEC-2000  
 LOCUS AF325532  
 DEFINITION Mus musculus heparin/heparan sulfate:glucuronic acid C5 epimerase  
 mRNA, complete cds.  
 ACCESSION AF325532  
 VERSION AF325532.1 GI:11935176  
 KEYWORDS  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 2088)  
 AUTHORS Crawford,B.E., Olson,S.K., Pinhal,M.A.S. and Esko,J.D.

TITLE Cloning, Golgi localization, and Enzyme Activity of the Full Length  
 Heparin/Heparan Sulfate:Glucuronic Acid C5 Epimerase  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 2088)  
 AUTHORS Crawford,B.E., Olson,S.K., Pinhal,M.A.S. and Esko,J.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-DEC-2000) Cellular and Molecular Medicine, University  
 of California, San Diego, 9500 Gilman Drive, La Jolla, CA  
 92093-0687, USA

FEATURES  
 source  
 1..2088  
 /organism="Mus musculus"  
 /strain="LAF1"  
 /db\_xref="taxon:10090"  
 /tissue\_type="Furth mastocytoma"  
 232..2088  
 /note="heparan sulfate/heparin biosynthesis"  
 /codon\_start=1  
 /product="heparin/heparan sulfate:glucuronic acid C5  
 epimerase"  
 /protein\_id="AA042004.1"  
 /db\_xref="GI:11935177"  
 /translation="MRCIAARVNRKTLIIICALETLVTLMLNCKSSDKAIOEPRLHS  
 GFRVDGLEKRSASSESNHVNHIKQSEBAPFOEQKAPVVGFGNSNGSKVLGL  
 KYEIDCLINDEHTIKGRREGNEVEFLPTWVEKYFDYKGVQYDGRFESYSK  
 VYAOEPSYHPDGVFMSPEGVNVEYRDRIKPCISGVEYPLSTOMGPOGYFPIAOYG  
 LSHYSKNLTERPHIYEYETADEDDRIIRPMEYVPGCPMAASADSRSTNKOFLA  
 PETSGVSIQGNKRDPIISDLKILNRSVSVLETERKQLETVHVSNTDILAR  
 DRDIYIGCPRTSMSTVTRDLVTDLRGKGVLSNTRKAVPKIMRKVRLIANGKPL  
 DNITISTAHNAAFNASDMLVRNODERGVNIVTRKLEGEFSLSPGYVSMAOQ  
 AISTLVRAVLLTKDYVLSALRATAYKFPSEQHGKAVAMNHDWYERYPTPSF  
 VLNGFMYSLIGLIDYKLETAETLEKARPLYERMESELSKMLPLDYDSGTYDLRIF  
 MGIAPNLAIRMDYHTTHINOLISTIDESIIFEFYKRNKSYLKSRAHN"

BASE COUNT 601 a 470 c 503 g 514 t  
 ORIGIN

Query Match 85.7%; Score 1203.2; DB 94; Length 2088;  
 Best Local Similarity 91.5%; Pred. No. 0;  
 Matches 1274; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

Qy 13 tctcaatagctatccaagctatgacagagagcccttcaacctgactgtgtt 72  
 Db 694 TCTCACTACTATTCGAAAGTATGACACAGATACCTTATCAACCTGACGGTGT 753  
 Qy 73 atgtccttgaaaggtcaaatgtaggaatccgagacagatgcatgaatgtgagtt 132  
 Db 754 ATGTCTTTGAAGGCTTAAATGTGAAGTCCGAGACAGATCAATGATTAAGTGAGTT 813  
 Qy 133 gaaggtgactcttctacacagtgaggacctcaagctcttcttcccaatccagat 192  
 Db 814 GAAGGTGCGCATATATCAACCAAGTGGGCGCTCAAGGCTATTTCTACCCATCAAT 873  
 Qy 193 gcaagatgaggttaagctactacagcaagaatctaaactgaaaaacccctcatataga 252  
 Db 874 GCACAGTATGGGCTAAGCTATTAACAGCAAGATTAACCCGGAACCCCTCATATGAA 933  
 Qy 253 gtatataaacaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 312  
 Db 934 GTATATGAACAGACAGAGACAGAGACAGAAACATCAAGACTTAATGATGACTGTGCC 993  
 Qy 313 aagggctgcttataagctagctgctgataagcaagattcaacaaatgtttaaagcttc 372  
 Db 994 AAGGGGTCTCTTCAATGGCCAGTGTGGACAGACAGCTATGATCCACCAATGTTAAAGCTTT 1053  
 Qy 373 attgtccagaagaacagatgaaggtgtatccttgcaacttgggggaacacaaagaattat 432  
 Db 1054 ATTGCTCCAGAAACCAATGAAGGTGTCTTTGCAAGTGGGAAACACAAAAAGCTTCAT 1113  
 Qy 433 attcatctgacctcaagctcttcaacaaatggaagcgtgctgtgtgttggagagaca 492  
 Db 1114 ATTTCATTTGACCTCAAGCTTTTAAACAAATGGAGTGTGTGTGTGTGTGTGTGTGTGTGT 1173

```

OY 493 gaaagaatcagctctcctacgtacatctatctcctcaaatccaccgaatgctctttaa 552
Db 1174 GAAAGAATCAGCTCTTCACTGATCTATGCTCAACACACGAGCTGATGCTTTTACA 1233
OY 553 gaaagaatcagctctcctacgtacatctcctcaaatccaccgaatgctctttaa 612
Db 1234 GAAAGAATCAGCTCTTCACTGATCTATGCTCAACACACGAGCTGATGCTTTTACA 1293
OY 613 gaaagaatcagctctcctacgtacatctcctcaaatccaccgaatgctctttaa 672
Db 1294 GAAAGAATCAGCTCTTCACTGATCTATGCTCAACACACGAGCTGATGCTTTTACA 1353
OY 673 gaaagaatcagctctcctacgtacatctcctcaaatccaccgaatgctctttaa 732
Db 1354 GAAAGAATCAGCTCTTCACTGATCTATGCTCAACACACGAGCTGATGCTTTTACA 1413
OY 733 gaaagaatcagctctcctacgtacatctcctcaaatccaccgaatgctctttaa 792
Db 1414 GAAAGAATCAGCTCTTCACTGATCTATGCTCAACACACGAGCTGATGCTTTTACA 1473
OY 793 gaaagaatcagctctcctacgtacatctcctcaaatccaccgaatgctctttaa 852
Db 1474 GAAAGAATCAGCTCTTCACTGATCTATGCTCAACACACGAGCTGATGCTTTTACA 1533
OY 853 gaaagaatcagctctcctacgtacatctcctcaaatccaccgaatgctctttaa 912
Db 1534 GAAAGAATCAGCTCTTCACTGATCTATGCTCAACACACGAGCTGATGCTTTTACA 1593
OY 913 gaaagaatcagctctcctacgtacatctcctcaaatccaccgaatgctctttaa 972
Db 1594 GAAAGAATCAGCTCTTCACTGATCTATGCTCAACACACGAGCTGATGCTTTTACA 1653
OY 973 gaaagaatcagctctcctacgtacatctcctcaaatccaccgaatgctctttaa 1032
Db 1654 GAAAGAATCAGCTCTTCACTGATCTATGCTCAACACACGAGCTGATGCTTTTACA 1713
OY 1033 gaaagaatcagctctcctacgtacatctcctcaaatccaccgaatgctctttaa 1092
Db 1714 GAAAGAATCAGCTCTTCACTGATCTATGCTCAACACACGAGCTGATGCTTTTACA 1773
OY 1093 gaaagaatcagctctcctacgtacatctcctcaaatccaccgaatgctctttaa 1152
Db 1774 GAAAGAATCAGCTCTTCACTGATCTATGCTCAACACACGAGCTGATGCTTTTACA 1833
OY 1153 gaaagaatcagctctcctacgtacatctcctcaaatccaccgaatgctctttaa 1212
Db 1834 GAAAGAATCAGCTCTTCACTGATCTATGCTCAACACACGAGCTGATGCTTTTACA 1893
OY 1213 gaaagaatcagctctcctacgtacatctcctcaaatccaccgaatgctctttaa 1272
Db 1894 GAAAGAATCAGCTCTTCACTGATCTATGCTCAACACACGAGCTGATGCTTTTACA 1953
OY 1273 gaaagaatcagctctcctacgtacatctcctcaaatccaccgaatgctctttaa 1332
Db 1954 GAAAGAATCAGCTCTTCACTGATCTATGCTCAACACACGAGCTGATGCTTTTACA 2013
OY 1333 gaaagaatcagctctcctacgtacatctcctcaaatccaccgaatgctctttaa 1392
Db 2014 GAAAGAATCAGCTCTTCACTGATCTATGCTCAACACACGAGCTGATGCTTTTACA 2073
OY 1393 gaaagaatcagctctcctacgtacatctcctcaaatccaccgaatgctctttaa 1452
Db 2074 GAAAGAATCAGCTCTTCACTGATCTATGCTCAACACACGAGCTGATGCTTTTACA 2113

```

```

RESULT 5
AC026992 167187 bp DNA HTG 03-JUL-2000
LOCUS Homo sapiens chromosome 15 clone RP11-352D13 map 15, WORKING DRAFT
DEFINITION
SEQUENCE, 30 unordered pieces.
AC026992
AC026992.4 GI:8901256
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

```

SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.  
AUTHORS 1 (bases 1 to 167187)  
TITLE Birren, B., Linton, L., Nusbaum, C., and Lander, E.  
JOURNAL Homo sapiens chromosome 15, clone RP11-352D13  
REFERENCE 2 (bases 1 to 167187)  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Baran, N., Bastien, V., Bedalov, F., Boguslavsky, L., Boudgalter, B., Brown, A., Burkett, G., Campiano, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domingo, M., Doyle, M., Fereira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Labocque, K., Lamazares, R., Landers, T., Lehotzky, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meldrum, J., Menus, L., Mihova, T., Miranda, C., Mieng, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Teste, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J., Vasiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A., and Zody, M.  
TITLE Direct Submission  
JOURNAL Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Jul 3, 2000 this sequence version replaced gi:8079836. All repeats were identified using RepeatMasker:  
Smt, A.F.A. & Green, P. (1996-1997)  
http://ffp.genome.washington.edu/RM/RepeatMasker.html  
Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIDR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
Project Information  
Center project name: L8712  
Center clone name: 352.D-13  
Summary Statistics  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 147074 bases at least Q40  
Consensus quality: 157353 bases at least Q30  
Consensus quality: 161662 bases at least Q20  
Insert size: 163000; agarose-fp  
Insert size: 164287; sum-of-coverage  
Quality coverage: 3.5 in Q20 bases; sum-of-coverage  
Quality coverage: 3.4 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently consists of 30 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1135: contig of 1135 bp in length  
1136 1235: gap of 100 bp  
1236 2789: contig of 1554 bp in length  
2790 2889: gap of 100 bp  
2890 4922: contig of 2033 bp in length  
4923 5022: gap of 100 bp  
5023 6180: contig of 1158 bp in length  
6181 6280: gap of 100 bp

```
* 6281 7551: contig of 1271 bp in length
* 7552 7651: gap of 100 bp
* 7652 9449: contig of 1798 bp in length
* 9450 9549: gap of 100 bp
* 9550 11032: contig of 1483 bp in length
* 11033 11132: gap of 100 bp
* 11133 12790: contig of 1658 bp in length
* 12791 12890: gap of 100 bp
* 12891 15198: contig of 2308 bp in length
* 15199 15298: gap of 100 bp
* 15299 17490: contig of 2192 bp in length
* 17491 17590: gap of 100 bp
* 17591 20019: contig of 2429 bp in length
* 20020 20119: gap of 100 bp
* 20120 23909: contig of 3790 bp in length
* 23910 24009: gap of 100 bp
* 24010 27099: contig of 3090 bp in length
* 27100 27199: gap of 100 bp
* 27200 30782: contig of 3583 bp in length
* 30783 30882: gap of 100 bp
* 30883 33932: contig of 3050 bp in length
* 33933 34032: gap of 100 bp
* 34033 37691: contig of 3659 bp in length
* 37692 37791: gap of 100 bp
* 37792 42856: contig of 5065 bp in length
* 42857 42956: gap of 100 bp
* 42957 49339: contig of 6333 bp in length
* 49340 49439: gap of 100 bp
* 49440 54849: contig of 5410 bp in length
* 54850 54949: gap of 100 bp
* 54950 58121: contig of 3172 bp in length
* 58122 58221: gap of 100 bp
* 58222 64249: contig of 6028 bp in length
* 64250 64349: gap of 100 bp
* 64350 70780: contig of 6431 bp in length
* 70781 70880: gap of 100 bp
* 70881 76628: contig of 5748 bp in length
* 76629 76728: gap of 100 bp
* 76729 85215: contig of 8487 bp in length
* 85216 85315: gap of 100 bp
* 85316 95800: contig of 10485 bp in length
* 95801 95900: gap of 100 bp
* 95901 105851: contig of 9951 bp in length
* 105852 105951: gap of 100 bp
* 105952 114860: contig of 8909 bp in length
* 114861 114960: gap of 100 bp
* 114961 123755: contig of 8795 bp in length
* 123756 123855: gap of 100 bp
* 123856 139927: contig of 16072 bp in length
* 139928 140027: gap of 100 bp
* 140028 167187: contig of 27160 bp in length.
FEATURES
source
1..167187
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-352D13"
/clone_lib="RPCT-II Human Male BAC"
1..1135
/misc_feature
/feature="assembly_fragment"
1236..2789
/misc_feature
/feature="assembly_fragment"
2890..4922
/misc_feature
/feature="assembly_fragment"
5023..6180
/misc_feature
/feature="assembly_fragment"
6281..7551
/misc_feature
/feature="assembly_fragment"
7652..9449
/misc_feature
/feature="assembly_fragment"
9550..11032
/misc_feature
/feature="assembly_fragment"
11033..12790
/misc_feature
/feature="assembly_fragment"
12791..15198
/misc_feature
/feature="assembly_fragment"
15199..17490
/misc_feature
/feature="assembly_fragment"
17491..20019
/misc_feature
/feature="assembly_fragment"
20020..23909
/misc_feature
/feature="assembly_fragment"
23910..27099
/misc_feature
/feature="assembly_fragment"
27100..30782
/misc_feature
/feature="assembly_fragment"
30783..33932
/misc_feature
/feature="assembly_fragment"
33933..34032
/misc_feature
/feature="assembly_fragment"
34033..37691
/misc_feature
/feature="assembly_fragment"
37692..42856
/misc_feature
/feature="assembly_fragment"
42857..42956
/misc_feature
/feature="assembly_fragment"
42957..49339
/misc_feature
/feature="assembly_fragment"
49340..49439
/misc_feature
/feature="assembly_fragment"
49440..54849
/misc_feature
/feature="assembly_fragment"
54850..54949
/misc_feature
/feature="assembly_fragment"
54950..58121
/misc_feature
/feature="assembly_fragment"
58122..58221
/misc_feature
/feature="assembly_fragment"
58222..64249
/misc_feature
/feature="assembly_fragment"
64250..64349
/misc_feature
/feature="assembly_fragment"
64350..70780
/misc_feature
/feature="assembly_fragment"
70781..76628
/misc_feature
/feature="assembly_fragment"
76629..76728
/misc_feature
/feature="assembly_fragment"
76729..85215
/misc_feature
/feature="assembly_fragment"
85216..95800
/misc_feature
/feature="assembly_fragment"
95801..95900
/misc_feature
/feature="assembly_fragment"
95901..105851
/misc_feature
/feature="assembly_fragment"
105852..105951
/misc_feature
/feature="assembly_fragment"
105952..114860
/misc_feature
/feature="assembly_fragment"
114861..114960
/misc_feature
/feature="assembly_fragment"
114961..123755
/misc_feature
/feature="assembly_fragment"
123756..123855
/misc_feature
/feature="assembly_fragment"
123856..139927
/misc_feature
/feature="assembly_fragment"
139928..140027
/misc_feature
/feature="assembly_fragment"
140028..167187
Query Match 66.5%: Score 933; DB 71; Length 167187;
Best Local Similarity 93.3%: Pred. No. 6,4e-245;
Matches 975; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
Qy 360 ttttaaaagcttcgtcgaagaacagtgagtgatctcttgaactggggaacac 419
Db 102759 TGTGGATATTCCTCCCTACAGAAACAGTGAAGGTGTATCTTGCACTGGGAAAC 102818
Qy 420 aaaagatttatttcaatttgacctgaagttcttcaaaatggaagcgtgtcgtg 479
Db 102819 AAAAGATTTATTTATTCATTGACCTCAAGTTCTTGCAAAATGGAAGTGTCCGTGT 102878
Qy 480 tctgagagcagcaaaagaatcagctcttcactgttcatatagtctcaaatcccaagct 539
Db 102879 TCTAGAGACACAGAAAAGAAATCAGCTTCATATACATTATGTCTCAAAATGCTCAGCT 102938
Qy 540 aattgctttaaagaagacataactatgcatcgggcccagaacatcagggagac 599
Db 102939 AATTGCTTTTAAAGAAAGATATATATCTATGCGATTTGGCCCAAGAACTTCATGGAGCAC 102998
Qy 600 agttaccgggacctgtcactgacctcaggaagagtggtcttttccacacaaagc 659
Db 102999 AGTTACCAGGAGCTGTGCTACTGACTCAGGAAAGAGGTGGTCTTTCAAAACAAAG 103058
```





```

FEATURES
source      Location/Qualifiers
1. 171073

```

BASE COUNT	55739	a	30189	c	29971	g	53068	t	2106	others
ORIGIN										
Query Match	34.9%; Score 490.6; DB 75; Length 171073;									
Best Local Similarity	94.6%; Pred. No. 1.6e-123;									
Matches 508;	Conservative 0; Mismatches 29; Indels 0; Gaps 0									
QY	868	ccagagtggtactcgcgcacatgagcccaagggcaagcattctacatgacagggcccat	927							
Db	4099	CCAGAGTGTATTTCTGCCATGGGCCAAGGCAAGCAATTTCTACATTATGTCAGGCGCTAT	4158							
QY	928	ctcttaacaaaacacatattcttcaatttaagctttaaaggaacagcccttacaag	987							
Db	4159	CTGTAAACAAAACCAATATATTCCTCAATTCAGCTTTTAAGGGCAACAGCCCTTATTAAG	4218							
QY	988	ttctgtcagagcagcatgagtcgaaggtctgtgttatacgaatacagaatgatatga	1047							
Db	4219	TTTCTATCTGAGCAGCATGAGTAAAGCTGTGTATGATTAACATGACTGTATGA	4278							
QY	1048	gaataatccaactacactatgctcttcttgtttaaagcttatagtatccttaattg	1107							
Db	4279	GAATATCCAAACCAACCTTGTCTTTTGTATTAATGCTTTATGTATCTTAAATATG	4338							
QY	1108	ctctatgaactaaagaacactgcaggggaaaaaacctcggaagaagaaggaggtctgt	1167							
Db	4339	CTGTATGACTTAAAGAAACCTGCAGGGGAAATCTGGAAAGAAAGCAAGGCTCTGTAT	4398							
QY	1168	gaagcgtgacatggaatcccttaaaagcatalgctccctctgtacagacactggtcaga	1227							
Db	4399	GAGCGTGGCATGGAATCTTTAAACCCATGCGCCCTGTATGACACTGGCTCAGGAAC	4458							
QY	1228	atctatgaactccggcactctatgcttggcaatggccccaacctggcccgctgga	1287							
Db	4459	ATCTATGACTCCGCTACTTTCATGCTTGGCANTGCTCTTAACCTGCTGCTGGACTAT	4518							
QY	1288	cacaccacccacatcaatcaatcagctgactgttgacacattgatagtatcccaacttc	1347							
Db	4519	CATATCCACCCACATCAATCATGTTGAGCTACTCAATGATGATGATGATCCCAAGCTTC	4578							
QY	1348	aaagaattgtcaagaagtgtagaagctacacttaaaagcagcgggcaaaagcacaac	1404							
Db	4579	AAAGAAATTTGTCAAGAGGTGAAAAAGCTTAAAGGCAAGCAGGCAAGACACAAAC	4635							
RESULT	7									
LOCUS	AC026992/c	167187	bp	DNA	HTG	03-JUL-2000				
DEFINITION	Homo sapiens chromosome 15 clone RP11-352D13 map 15, WORKING DRAFT									
ACCESSION	AC026992	SEQUENCE, 30 unordered pieces.								
VERSION	AC026992.4	GI:8901256								
KEYWORDS	HTG; HTGS_PHAISEL; HTGS_DRAFT.									
SOURCE	human.									
ORGANISM	Homo sapiens									
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.									
TITLE	1 (bases 1 to 167187)									
JOURNAL	Birren, B., Linton, L., Nusbaum, C. and Lander, E.									
REFERENCE	Unpublished									
AUTHORS	2 (bases 1 to 167187)									

TITLE  
JOURNAL  
COMMENT

Klein, J., Larocque, K., Lamazares, R., Landers, T., Lebecky, J.,  
Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, R.,  
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,  
Meldrum, J., Meneu, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,  
Pisanic, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Testaferr, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 3, 2000 this sequence version replaced g1:8079836.  
All repeats were identified using RepeatMasker:  
Solt, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

## Project Information

Center project name: L8712

Center clone name: 352\_D\_13

## Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator BigDye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 147074 bases at least Q40

Consensus quality: 157353 bases at least Q30

Consensus quality: 161662 bases at least Q20

Insert size: 163000; agarose-1p

Insert size: 164287; sum-of-contents

Quality coverage: 3.5 in Q20 bases; agarose-1p

Quality coverage: 3.4 in Q20 bases; sum-of-contents

NOTE: This is a 'working draft' sequence. It currently  
consists of 30 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

```

1      1135: contig of 1135 bp in length
*      1136 1235: gap of 100 bp
*      1236 2789: contig of 1554 bp in length
*      2780 2889: gap of 100 bp
*      2880 4922: contig of 2033 bp in length
*      4923 5022: gap of 100 bp
*      5023 6180: contig of 1158 bp in length
*      6181 6280: gap of 100 bp
*      6281 7551: contig of 1271 bp in length
*      7552 7651: gap of 100 bp
*      7652 9449: contig of 1798 bp in length
*      9450 9549: gap of 100 bp
*      9550 11032: contig of 1483 bp in length
*      11033 11132: gap of 100 bp
*      11133 12790: contig of 1658 bp in length
*      12791 12890: gap of 100 bp
*      12881 15198: contig of 2308 bp in length
*      15199 15298: gap of 100 bp
*      15299 17490: contig of 2192 bp in length
*      17491 17590: gap of 100 bp
*      17591 20019: contig of 2429 bp in length
*      20020 20119: gap of 100 bp
*      20120 23909: contig of 3790 bp in length
*      23910 24009: gap of 100 bp
*      24010 27099: contig of 3090 bp in length
*      27100 27199: gap of 100 bp

```

## FEATURES

## source

```

*      27200 30782: contig of 3583 bp in length
*      30783 30883: gap of 100 bp
*      30883 33332: contig of 3050 bp in length
*      33933 34032: gap of 100 bp
*      34033 37691: contig of 3659 bp in length
*      37692 37791: gap of 100 bp
*      37792 42856: contig of 5065 bp in length
*      42857 42956: gap of 100 bp
*      42957 49333: contig of 6383 bp in length
*      49340 49439: gap of 100 bp
*      49440 54849: contig of 5410 bp in length
*      54850 54949: gap of 100 bp
*      54950 58121: contig of 3172 bp in length
*      58122 58221: gap of 100 bp
*      58222 64249: contig of 6028 bp in length
*      64250 64349: gap of 100 bp
*      64350 70780: contig of 6431 bp in length
*      70781 70880: gap of 100 bp
*      70881 76628: contig of 5748 bp in length
*      76629 76728: gap of 100 bp
*      76729 85215: contig of 8487 bp in length
*      85216 85315: gap of 100 bp
*      85316 95800: contig of 10485 bp in length
*      95801 95900: gap of 100 bp
*      95901 105851: contig of 9951 bp in length
*      105852 105951: gap of 100 bp
*      105952 114860: contig of 8909 bp in length
*      114861 114960: gap of 100 bp
*      114961 123755: contig of 8795 bp in length
*      123756 123855: gap of 100 bp
*      123856 139927: contig of 16072 bp in length
*      139928 140027: gap of 100 bp
*      140028 167187: contig of 27160 bp in length.

```

location/Qualifiers

```

1..167187
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-352D13"
/clone_11b="RP11 Human Male BAC"
1..1135
/feature="assembly-fragment"
1236..2789
/feature="assembly-fragment"
2780..2889
/feature="assembly-fragment"
2880..4922
/feature="assembly-fragment"
4923..5022
/feature="assembly-fragment"
5023..6180
/feature="assembly-fragment"
6181..6280
/feature="assembly-fragment"
6281..7551
/feature="assembly-fragment"
7552..7651
/feature="assembly-fragment"
7652..9449
/feature="assembly-fragment"
9450..9549
/feature="assembly-fragment"
9550..11032
/feature="assembly-fragment"
11033..11132
/feature="assembly-fragment"
11133..12790
/feature="assembly-fragment"
12791..12890
/feature="assembly-fragment"
12881..15198
/feature="assembly-fragment"
15199..15298
/feature="assembly-fragment"
15299..17490
/feature="assembly-fragment"
17491..17590
/feature="assembly-fragment"
17591..20019
/feature="assembly-fragment"
20020..20119
/feature="assembly-fragment"
20120..23909
/feature="assembly-fragment"
23910..24009
/feature="assembly-fragment"
24010..27099
/feature="assembly-fragment"
27100..27199
/feature="assembly-fragment"

```



Query Match	14.58;	Score 204;	DB 71;	Length 167187;
Best Local Similarity	90.98;	Pred. No. 6.8e-45;		
Matches 229; Conservative	0;	Mismatches 20;	Indels 3;	Gaps 1

	RESULT						
	8						
LOCUS	AC018117/c						
DEFINITION	Ac018117	29372 bp	DNA	HTG	09-DEC-1999		
	<i>Drosophila melanogaster</i> , *** SEQUENCING IN PROGRESS ***, in ordered pieces						
ACCESSION	AC018117						
VERSION	AC018117.1	GI:6553074					
KEYWORDS	HTG; HTGS_PHASE2.						
SOURCE	fruit fly.						
COMMENTARY							

Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive

Query Match	10.7%;	Score 150.8;	DB 65;	Length 29372;
Best Local Similarity	63.0%;	Pred. No. 2.6e-30;		
Matches 233;	Conservative 0;	Mismatches 137;	Indels 0;	Gaps 0;

RESULT	9
AC007593/c	
LOCUS	AC007593 119597 bp DNA HTG 21-FEB-2000
DEFINITION	Drosophila melanogaster chromosome 2 clone BACR01C10 (ID650) RPc1-98
FEATURES	01..C10 map 42E-43A strain Y: cn bw sp, *** SEQUENCING IN PROGRESS
CONNECTION	*** 89 unordered pieces.
COMMENT	0007593

ACCESSION	AC007593
VERSION	AC007593.4
KEYWORDS	GI:7018749
SOURCE	HTG: HTGS_PHASE1
ORGANISM	fruit fly. <i>Drosophila melanogaster</i> Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroptera; drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 119597)
AUTHORS	Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Bultmann, C., Champe, M., Chavez, C., Chew, M., Chishti, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hinkle, A., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Lee, B., Lewis, S., Li, P., Ling, H., Mosher, A.A., Mosher, M., Nixon, R., Paclob, J.M., Park, S., Pfeiffer, B., Richards, S., Sethi, H., Svirskas, R.R., Wan, K.H., Webster, D., Woolley, P., Yang, S., Yee, M., Yu, C. and Rubin, G.M.
TITLE	Sequencing of <i>Drosophila melanogaster</i>
JOURNAL	Unpublished

REFERENCE  
AUTHORS  
2 (bases 1 to 119597)  
Celikler,S.E., Abayani,A., Arcaina,T.T., Baxter,E., Blazek,R.G.,  
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,  
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,R., Kearney,L.,  
Kim,E., Lee,B., Lewis,S., Li,P., Lomutan,M.A., Mazda,P.,  
Moshirefi,A.R., Moshirefi,M., Nixon,K., Peckel,J.M., Park,S.,  
Pfeiffer,B., Poon,L., Segueira,A., Sethi,H., Sult,E.,  
Svitskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
Rubin,G.M.  
Direct Submission  
Submitted (20-MAY-1999) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA g1:4885670.  
On Feb 21, 2000 this sequence version replaced g1:4885670.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bugreport@fruitfly.berkeley.edu](mailto:bugreport@fruitfly.berkeley.edu). All contigs in this submission meet  
the following cutoffs: length >= 200 bases.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 89 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 584: contig of 584 bp in length  
\* 585 664: gap of unknown length  
\* 665 1363: contig of 699 bp in length  
\* 1364 1443: gap of unknown length  
\* 1444 2281: contig of 738 bp in length  
\* 2182 2262: gap of unknown length  
\* 2262 3104: contig of 843 bp in length  
\* 3105 3184: gap of unknown length  
\* 3185 3906: contig of 722 bp in length  
\* 3907 3986: gap of unknown length  
\* 3987 4920: contig of 934 bp in length  
\* 4921 5000: gap of unknown length  
\* 5001 5966: contig of 966 bp in length  
\* 5967 6046: gap of unknown length  
\* 6047 6862: contig of 816 bp in length  
\* 6863 6942: gap of unknown length  
\* 6943 7461: contig of 519 bp in length  
\* 7462 7541: gap of unknown length  
\* 7542 8246: contig of 705 bp in length  
\* 8247 8326: gap of unknown length  
\* 8327 9098: contig of 772 bp in length  
\* 9099 9178: gap of unknown length  
\* 9179 9852: contig of 674 bp in length  
\* 9853 9932: gap of unknown length  
\* 9933 10800: contig of 868 bp in length  
\* 10801 10880: gap of unknown length  
\* 10881 12098: contig of 1218 bp in length  
\* 12099 12178: gap of unknown length  
\* 12179 13330: contig of 1152 bp in length  
\* 13331 13410: gap of unknown length  
\* 13411 14082: contig of 672 bp in length  
\* 14083 14162: gap of unknown length  
\* 14163 14710: contig of 548 bp in length  
\* 14711 14790: gap of unknown length  
\* 14791 15379: contig of 589 bp in length  
\* 15380 15459: gap of unknown length  
\* 15460 16354: contig of 895 bp in length  
\* 16355 16434: gap of unknown length  
\* 16435 17123: contig of 689 bp in length  
\* 17124 17203: gap of unknown length  
\* 17204 18522: contig of 1319 bp in length  
\* 18523 18602: gap of unknown length  
\* 18603 19730: contig of 1128 bp in length  
\* 19731 19810: gap of unknown length  
\* 19811 20869: contig of 1059 bp in length  
\* 20870 20949: gap of unknown length

\* 20950 22525: contig of 1576 bp in length  
\* 22526 22605: gap of unknown length  
\* 22606 23882: contig of 1277 bp in length  
\* 23883 23962: gap of unknown length  
\* 23963 25201: contig of 1239 bp in length  
\* 25202 25281: gap of unknown length  
\* 25282 26311: contig of 1030 bp in length  
\* 26312 26391: gap of unknown length  
\* 26392 27304: contig of 913 bp in length  
\* 27305 27384: gap of unknown length  
\* 27385 28957: contig of 1573 bp in length  
\* 28958 29037: gap of unknown length  
\* 29038 29951: contig of 914 bp in length  
\* 29952 30031: gap of unknown length  
\* 30032 30952: contig of 921 bp in length  
\* 30953 31032: gap of unknown length  
\* 31033 31672: contig of 640 bp in length  
\* 31673 31752: gap of unknown length  
\* 31753 32764: contig of 1012 bp in length  
\* 32765 32844: gap of unknown length  
\* 32845 34084: contig of 1240 bp in length  
\* 34085 34164: gap of unknown length  
\* 34165 35006: contig of 842 bp in length  
\* 35007 35086: gap of unknown length  
\* 35087 36296: contig of 1210 bp in length  
\* 36297 36376: gap of unknown length  
\* 36377 37977: contig of 1601 bp in length  
\* 37978 38057: gap of unknown length  
\* 38058 39491: contig of 1434 bp in length  
\* 39492 39571: gap of unknown length  
\* 39572 40645: contig of 1074 bp in length  
\* 40646 40725: gap of unknown length  
\* 40726 42319: contig of 1594 bp in length  
\* 42320 42399: gap of unknown length  
\* 42400 43547: contig of 1148 bp in length  
\* 43548 43627: gap of unknown length  
\* 43628 45038: contig of 1411 bp in length  
\* 45039 45118: gap of unknown length  
\* 45119 46182: contig of 1064 bp in length  
\* 46183 46262: gap of unknown length  
\* 46263 48155: contig of 1893 bp in length  
\* 48156 48235: gap of unknown length  
\* 48236 49922: contig of 1687 bp in length  
\* 49923 50002: gap of unknown length  
\* 50003 51972: contig of 1970 bp in length  
\* 51973 52052: gap of unknown length  
\* 52053 54342: contig of 2290 bp in length  
\* 54343 54422: gap of unknown length  
\* 54423 55950: contig of 1528 bp in length  
\* 55951 56030: gap of unknown length  
\* 56031 57106: contig of 1076 bp in length  
\* 57107 57186: gap of unknown length  
\* 57187 58601: contig of 1415 bp in length  
\* 58602 58681: gap of unknown length  
\* 58682 59984: contig of 1303 bp in length  
\* 59985 60064: gap of unknown length  
\* 60065 61596: contig of 1532 bp in length  
\* 61597 61676: gap of unknown length  
\* 61677 63106: contig of 1430 bp in length  
\* 63107 63186: gap of unknown length  
\* 63187 65066: contig of 1880 bp in length  
\* 65067 65146: gap of unknown length  
\* 65147 67491: contig of 2345 bp in length  
\* 67492 67571: gap of unknown length  
\* 67572 71138: contig of 3567 bp in length  
\* 71139 71218: gap of unknown length  
\* 71219 73245: contig of 2027 bp in length  
\* 73246 73325: gap of unknown length  
\* 73326 76254: contig of 2929 bp in length  
\* 76255 76334: gap of unknown length  
\* 76335 79184: contig of 2850 bp in length  
\* 79185 79264: gap of unknown length  
\* 79265 83109: contig of 3845 bp in length



```

* 40574 40553: gap of unknown length
* 40654 50146: contig of 9493 bp in length
* 50147 50226: gap of unknown length
* 50227 56630: contig of 6404 bp in length
* 56631 56710: gap of unknown length
* 56711 65243: contig of 8533 bp in length
* 65244 65323: gap of unknown length
* 65324 78848: contig of 13525 bp in length
* 78849 78928: gap of unknown length
* 78929 106751: contig of 27823 bp in length
* 106752 130045: contig of 23214 bp in length
* 130046 130125: contig of unknown length
* 130126 163293: contig of 35168 bp in length
* 163294 165373: gap of unknown length
* 165374 165599: contig of 226 bp in length
* 165600 165679: gap of unknown length
* 165680 166287: contig of 608 bp in length
* 166288 167062: gap of unknown length
* 167062 167062: contig of 695 bp in length.

```

```

FEATURES
  source
    1..167062
      /organism="Drosophila melanogaster"
      /strain="Y: cn bw sp"
      /db_xref="taxon:7227"
      /chromosome="2"
      /map="42E-43A"
      /clone="BACR10F15 (D621) RPCT-98 10.F.15"
      /clone_lib="RPCT-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
PBAC63.6"

```

```

BASE COUNT      45406 a 36877 c 36482 g 46377 t 1920 others
ORIGIN

```

```

Query Match      10.7%; Score 150.8; DB 60; Length 167062;
Best Local Similarity 63.0%; Pred. No. 2.7e-30;
Matches 233; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

```

```

OY 1035 tgaactgtatgaagaataccaaactaactaactgtctttgttttaagtgtttatga 1094
Db 153555 TTACAGCTACGAGAGATACCCACACGACCCCTTATGCTTAAACGGCTTTATCTA 153496
OY 1095 tttcttaattggcgctgctgactaactaagaactcagaggaactcggagaaagc 1154
Db 153495 TTCTACTGCTGGGTCTTACATCAACACACTGCCCCGGGAAATAATGCCGGGAGGC 153436
OY 1155 gaagtccttgaatgagcgtgagatgaatcccttaagcactgctcccttgaagcac 1214
Db 153435 TGAAGAGCTTTTCGCGAGGCGATGCACTCCCTTAAATAAGATGCTTTGATTTGACAC 153376
OY 1215 tggctcaggaacacatctatgacctcggcgaacttcattgcttggaatgcccccaactgagc 1274
Db 153375 TGGCTCCGGGCGACGATGACGCTGCGCCATCTGAGCTGGGCTACCTCTTAATCTGGC 153316
OY 1275 ccgctgggaactatacaaccacacacatcaactcgaactgactgcttaagccacttga 1334
Db 153315 GCGCTGGGACACTGACATGACGACGTAACCACTTCTTACCAACATCGACAG 153256
OY 1335 gtccccaactctcaagaattgtcaagaagtggaagacttaccttaagggcagcgagc 1394
Db 153255 CGATTCATTATTCACACAACTCGGAGCGCTGGAAGGCTATATGTTGCTCGGGC 153196
OY 1395 aaagcacac 1404
Db 153195 CAAGCACAC 153186

```

```

RESULT 11
LOCUS      AE003789 299474 bp DNA INV 04-OCT-2000
DEFINITION Drosophila melanogaster genomic scaffold 14200001336042 section 1
of 2, complete sequence.

```

```

ACCESSION      AE003789 AE002778
VERSION        AE003789.2 GI:107272129
KEYWORDS
SOURCE
ORGANISM
REFERENCE
  AUTHORS
    1 (bases 1 to 299474)
    Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
    Ananidis,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
    George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
    Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
    Brandon,R.C., Rogers,Y.H., Blazer,V.G., Chame,M., Pfeiffer,B.D.,
    Man,K.H., Doyle,C., Baxter,E.G., Helt,G., Helt,G., Nelson,C.R., Gabor
    Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
    Andrews-Pfannkoch,C., Bateman,D., Ballew,R.M., Basu,A.,
    Baxendale,J., Bayraktaroglu,U., Beasley,E.M., Beeson,K.Y.,
    Benos,P.V., Bertram,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
    Botchan,M.R., Bouck,J., Brokstein,P., Brothier,P., Burtis,K.C.,
    Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
    Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
    Pablo,B., Delcher,A., Deng,Z., Dey,A.D., Dew,I., Dietz,S.M.,
    Dodson,K., Dou,P.L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
    Dunn,P., Durbin,K.J., Evangelista,C.C., Ferriz,C., Ferriter,S.,
    Fleischmann,W., Foster,C., Gabriellian,A.E., Garg,N.S.,
    Gelbart,W.M., Glasser,K., Glodde,A., Gong,F., Gorrell,J.H., Gu,Z.,
    Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
    Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
    Wei,M.H., Idagawa,C., Jallil,M., Kalush,F., Karpen,G.H., Ke,Z.,
    Kennison,J.A., Ketchum,K.A., Kimmel,P.E., Kodira,C.D., Kraft,C.,
    Kravitz,S., Kulp,D., Lai,Z., Laake,P., Lei,Y., Levitsky,A.A.,
    Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mathe,B., McIntosh,T.C.,
    McLeod,M.P., McPherson,D., Merkulov,G., Mitsuhashi,N.V., Mobarry,C.,
    Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
    Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
    Nuskern,D.R., Paclab,J.M., Palazolo,M., Pittman,G.S., Pan,S.,
    Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
    Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
    Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
    Stapleton,M., Strong,R., Sun,E., Svrskas,R., Tector,C., Turner,R.,
    Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasarman,D.A.,
    Weinstein,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
    Wortley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
    Zaveri,J.S., Zhong,W., Zhou,G., Zhao,Q., Zheng,L., Zheng,X.H.,
    Zhong,F.N., Zhou,M., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
    Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
    The genome sequence of Drosophila melanogaster
    Science 287 (5461), 2185-2195 (2000)
  JOURNAL
  MEDLINE
  REFERENCE
  AUTHORS
  TITLE
  JOURNAL
  COMMENT
  FEATURES
    source
      1..299474
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
        /chromosome="2R"
        /join(11901..12167,12272..12335,12376..13327,37670..38167,
        38262..38332)
        /gene="1.28"
        /note="Nucleotide sequence of the Celera sequence differs
        from the published sequence for this transcript."
        /product="CP9093"
        /db_xref="FLYBASE:FBgn0009397"
        /db_xref="FLYBASE:FBgn0010347"
        /gene="1.28"
        /note="CG9397"
        /map="42A13-42A16"

```

gene



Db 180045 GCCTGTGGACTACATGCAACGACGTGACAGTACTTCTGTAGCCACATGACAG 179986  
 QY 1335 gtcccaacttcaagaatttcgaagagtggaagagctccctaaagcagccgagc 1394  
 Db 179985 CGATCCATTAATGTCACAAACGACGCGCTGGAAGGCGCTATCTTGTGCTGCGGSC 179926  
 QY 1395 aaagcacaac 1404  
 Db 179925 CAAGCACAAAC 179916

RESULT 12  
 CEB0285 41397 bp DNA INV 23-OCT-2000  
 LOCUS Caenorhabditis elegans cosmid B0285, complete sequence.  
 DEFINITION 234533  
 ACCESSION 234533  
 VERSION 1  
 KEYWORDS HMG; Aminoacyl-tRNA synthetase; CDG2-like protein kinase; Choline kinase; Glycine-rich domain; Muclin; transfer RNA; tRNA-Leu.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.  
 REFERENCE 1 (bases 1 to 41397)  
 AUTHORS none  
 TITLE Genome sequence of the nematode C. elegans: a platform for Investigating Biology. The C. elegans Sequencing Consortium  
 JOURNAL Science 282 (5396), 2012-2018 (1998)  
 MEDLINE 99069613  
 REMARK The C. elegans Sequencing Consortium.  
 Erratum: [[published errata appear in Science 1999 Jan 1;283(5398):335 and 1999 Mar 26;283(5410):2103 and 1999 Sep 3;285(5433):1493]]  
 2 (bases 1 to 41397)  
 REFERENCE Direct Submission  
 AUTHORS Sulston, J.E.  
 TITLE Submitted (25-JUN-1994) Nematode Sequencing Project, Sanger Centre, Hinxton, Cambridge CB10 1RQ, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or twenematode.wustl.edu  
 JOURNAL Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.  
 COMMENT Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.  
 IMPORTANT: This sequence is not the entire insert of clone B0285. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.  
 The true left end of clone B0285 is at 25507 in sequence AL024499.  
 The start of this sequence (1..104) overlaps with the end of sequence AL024499.  
 The end of this sequence (41294..41397) overlaps with the start of sequence Z30973.  
 Identical proteins B0285.8, B0285.9 and B0285.10 are 60-70% identical at the amino acid level.  
 For a graphical representation of this sequence and its analysis see: <http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=B0285>  
 [000905 dl] Merged B0285.1 and B0285.2 based on EST data.  
 location/Qualifiers  
 1..41397  
 /organism="Caenorhabditis elegans"  
 /db\_xref="taxon:6239"  
 /chromosome="III"  
 /clone="B0285"  
 join(2715..2908,3445..3642,5546..5672,6107..6220,6278..6498,6852..6997,7050..7339,7398..7471,7733..7888,

8139..8333,8382..8532,8827..9015)  
 /gene="B0285.1"  
 join(2715..2908,3445..3642,5546..5672,6107..6220,6278..6498,6852..6997,7050..7339,7398..7471,7733..7888,8139..8333,8382..8532,8827..9015)  
 /gene="B0285.1"  
 /note="contains similarity to Pfam domain: PF00069 (Eukaryotic protein kinase domain), Score=302.4, E-value=1.7e-87, N=1"  
 /codon\_start=1  
 /protein\_id="CAAB4302.2"  
 /db\_xref="GI:11024555"  
 /db\_xref="SWISS-PROT:P46551"  
 /translation="MEISPGSTHERKSGYGRRTSHSGSPRFRYSKDKRSGSR OGVRPRDRKDISIPQYKORNRSGGCGGRGNDSEFRKGRKGRKRSRSRSRSPKRSKSGSKSRKRSRSGSSSDMLSGELKKGHDGSSSKSRKRSRSHSSSSSSSGEADLPVSSNGAMVTAIPPPPPFNINPQPMSPDPPPPPPMTPPPRPPPPFIPSPVDIHPATASFSLSTIPPPPTDGGASSKRODPLPMPD SKRIATRPVITTRKGNATNRSDSDSWYKTNLHYTMLODIGGTGYQKAVNNLITG EOVALKRVLRENEKEGPITAIKEIKILROLHKNIVRLMDIYIDIDISDELKRTAN FYIYFEVDHDLGILEKELVDNKOICSLFROLGLAYIHTGFLHRLDKSNI LVNKGELKTDADIGLARIWKESRLVNRVITLTYRPPETELVGNTPRMKRTYORH MGLFLTRKPLPKNNEFCOLIELSKYSGSPANDPELTELVDGPAIDVWSYGC REFEHIMPREAVDLDMKTLFNEKTSKKEALNHPMITSLEHTYQPLKLDHDDC HEWMSKROKKSARLRQAESSGSGHSIRATSPHKGAGDELEIKISDITFQSENK VN"  
 complement(join(11398..11451,11504..11608,12102..12397,12467..12512))  
 /gene="B0285.3"  
 complement(join(11398..11451,11504..11608,12102..12397,12467..12512))  
 /gene="B0285.3"  
 /note="contains glycine-rich domain  
 cDNA EST YK530f5.3 comes from this gene"  
 /codon\_start=1  
 /protein\_id="CAAB4296.1"  
 /db\_xref="GI:3873727"  
 /db\_xref="SWISS-PROT:P46553"  
 /translation="MNYGNGGCGQOPRGNNMRPQGRGGCGGKGNDRGGRGGS SNDGSMNFRDGNLNFASPYQNRGNFRGGCGGCGGQNNYGNRGRGGRGGR GRGGCGGKNGFPIIDANKCIIPGMNPNDALEKEVEGVATFETKVPDPSPSS ABETTA"  
 join(12779..12919,12968..13145,13248..13412,13465..13603,13901..14107,14164..14250,14306..14390)  
 /gene="B0285.4"  
 join(12779..12919,12968..13145,13248..13412,13465..13603,13901..14107,14164..14250,14306..14390)  
 /gene="B0285.4"  
 /note="cDNA EST YK173g10.5 comes from this gene  
 cDNA EST YK173g10.3 comes from this gene  
 cDNA EST YK217f7.3 comes from this gene  
 cDNA EST YK217f7.5 comes from this gene"  
 /codon\_start=1  
 /protein\_id="CAAB4295.1"  
 /db\_xref="GI:3873726"  
 /db\_xref="SWISS-PROT:P46554"  
 /translation="MDEAVSSDSHVAIAITRRRSNSVSDSYVQRTNDADQCKF ATOKGKDEIFSRANSSSNVSEARPEISMYKMTAIEKYVDFLEDFGNQ VYSGGDFDLFWLVSSGAKLVKVEYDFSVSKIRILIKIGNSVDKSPES DAVYSHADLHAGNYHLICADLROANLDOKALCOLSHDIPFLIEACVLYWSAS STALKQIVSOPKOPAPVNYQCFRSDAFYVMONICDRGIDLOHGLEMCASARIE RFRNGFEVKYMDNOLFNNFLDQKEYSRLREIMDEMLLOQLFRAHYVSARIE join(14571..14615,14657..14818,15191..15485,15715..15810,15860..16018,16144..16259,16388..16658,17070..17288,17776..18311)  
 /gene="B0285.5"  
 join(14571..14615,14657..14818,15191..15485,15715..15810,15860..16018,16144..16259,16388..16658,17070..17288,17776..18311)  
 /note="cDNA EST YK96a2.5 comes from this gene"  
 /codon\_start=1  
 /protein\_id="CAAB4297.1"

CDS

gene

CDS

gene

CDS

gene

CDS

```

/dbd_xref="GI:3873728"
/translation="NWVLSIKRPNTISLAKPMKCLRRSRNRHIIYLLVACALFLINR
LQESRLEDEDEDELQDVNVEDKRIIEEPSPESIECIILANGKSMKCMWDEEDYF
LVYLSIKRFDQTLKGLDQVNEDEKFLYVYAKMSPESTYDLPFGHFTSYLVEEDYF
RCVSAQDLPVSTQMDPIRYUYDILQISQGLQYISMKIDLSINSEKSPKDDVLYF
NSKMKGAAGNHTTELEPIFYDDDOMKQVNIISGALALAAQYVLLSDPLHLYSPD
MKPEANSFVFLAKMKQDQLLVLYNVYSEGGKQSGDKRRKDDVYLAHADLVLSGFR
GELYVKKQITDRRQSHAFYAAADLVKQNDRGWSYFVRSIAERKLVLPWHS
AMAGHGELVLTAFRKEDEKYLKSAARALKFTKINSDDGVREFFGNLYEYPLT
TPGSEVNGELFYLIGLYDLSLKLALIDENDELMRAKIDQAEQLSACVSRKQLLPL
YDTCSGTITDRLHVALGTAPNLRAWYHAAVHYLLKWIIGIEKDEVLKTDPMRTGYA
YGRKRAK"
/length=21833.
21850.0,22021.
22170.22213.
22338.22611.
22646.
22695.
22853.22802.
22945.22990.
23098.23150.
23377.
23585.
23781.23825.
23946.23993.
24295.25051.
25206
/genes="B0285.6"
/length=21833.
21850.0,22021.
22170.22213.
22338.22611.
22646.
22695.
22853.22802.
22945.22990.
23098.23150.
23377.
23585.
23781.23825.
23946.23993.
24295.25051.
25206
/notes="similar to C. elegans R107.1, K08E5.2: localized
similarly to yeast URF YCR37C and rat mucin (PIR:S26784),
contains similarity to Pfam domain: Pf00939
(Sodium:sulfate symporter transmembrane region),
score=153.0, E-value=2.4e-43, N=2"
/codon_start=1
/protein_id="CAAB4299.1"
/db_xref="GI:3873730"
/db_xref="SWISS-PROT:P46556"
/translation="MPRESFEGKCAVYELILMTCGYAEVYPLAVTSFTPLMALPPLGI
VLSIEVAKTPEADTNVYEFNSMLSAVBECCQHKRIALAKMLTYGTPRHMLMGFMI
LITSTLSIMISTQACCAACMAIALALEEIMTPMRKEENKNI EYVKLTEDDEDEKE
KKKIDLSLSTVSRGRIKCAALMLVAHSLIGGIGITNSGPIIFRDMILENPNEDH
GISYLSMAAFELIPMLIYPMSSMYVLOLDFPRHLMGKEREPTEKOEBEVRAV
WKSVDGLGPMFAKSLVFLVFLAVLSYSDPKYIPGMSDIFRKGYYTDSGCLVY
FLILTPMKKKDPRPIPKKRSRVSQEDPLDMDCYARRFPMSITILLGAGFAISDAY
RVSGLSLTIACSLSTISKRPFFVMQIILSIYVVWTEFTSTMSATASFTIPISMAE
AVGHPHSTFPLPAIGSFSEFPLMPTLPAPNALVYEFKTRITMDVMSGCVPLNIFICIAI
TALNNMTWAFMLFMNGYFDPYALRAHNTMGTNSGOF"
/length=26002.
26439.9,27474.4.
27738.30045.
30685.31287.
31397.
32550.
32792.32888.
33229.33357.
33515.33765.
33911
/genes="B0285.7"
/length=26002.
26439.9,27474.4.
27738.30045.
30685.31287.
31397.
32550.
32792.32888.
33229.33357.
33515.33765.
33911
/notes="threonine-rich domain at N terminus: short region
of similarity with aminopeptidases, contains similarity to
Pfam domain: PF01443 (peptidase family M1), score=130.2,
E-value=1.6e-05, N=1
CDNA EST EMBL:U00449 comes from this gene
CDNA EST EMBL:T02449 comes from this gene
CDNA EST EMBL:T00914 comes from this gene
CDNA EST EMBL:T02161 comes from this gene
CDNA EST YK5106.5 comes from this gene
CDNA EST YK5106.3 comes from this gene
CDNA EST EMBL:JN75802 comes from this gene
CDNA EST YK48548.3 comes from this gene
CDNA EST YK48548.5 comes from this gene
CDNA EST YK327e7.3 comes from this gene
CDNA EST YK327e7.5 comes from this gene
CDNA EST YK345e7.5 comes from this gene
CDNA EST YK345a2.5 comes from this gene
CDNA EST YK266f2.5 comes from this gene
CDNA EST YK246f6.5 comes from this gene
CDNA EST YK230f6.3 comes from this gene

```

Query Match	5.48;	Score 75.8;	DB 5;	Length 41397;
Best Local Similarity	52.68;	Pred. No. 9.9e-10;		
Matches 210; Conservative	0;	Mismatches 162;	Indels 27;	Gaps 1;

[illegible]

RESULT	13
LOCUS	AL160170/c
DEFINITION	AL160170 156257 bp DNA HTG 20-JAN-2001 Homo sapiens chromosome 1 clone RP11-234F20, *** SEQUENCING IN PROGRESS ***, 16 unordered pieces.
ACCESSION	AL160170
VERSION	AL160170.4 GI:9863586
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	Plumb,B.
TITLE	Direct Submission
JOURNAL	Submitted (19-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT	On Aug 21, 2000 this sequence version replaced gi:9187197.

```

Web site: http://www.sanger.ac.uk
Contact: humgvery@sanger.ac.uk
-----
Project Information
Center project name: ba234fz0
-----
Summary Statistics
Assembly program: XGAP4, version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 147428 bases at least Q40
Consensus quality: 150614 bases at least Q30
Consensus quality: 152462 bases at least Q20
Insert size: 154757; sum-of-contigs
Insert size: 151975; 15.3% error; agarose-tp
Quality coverage: 3.45x in Q20 bases; sum-of-contigs quality
coverage: 3.61x in Q20 bases; agarose-tp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```







numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at  
<http://www.sanger.ac.uk/projects/C.elegans/wormpep> This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
Group. Further information can be found at  
<http://www.sanger.ac.uk/HEP/Chr1>

RP11-183M13 is from the library RP11-11.1 constructed at the  
Roswell Park Cancer Institute by the group of Pieter de Jong. For  
further details see <http://Dacpac.med.buffalo.edu/>  
VECTOR: pBAC3.6

This sequence is the entire insert of clone RP11-183M13 The true  
left end of clone RPS-831021 is at 111240 in this sequence.

## FEATURES

source  
1.167721  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-183M13"  
/clone\_id="RP11-11.1"  
166..337  
/note="L2 repeat: matches 2017..2195 of consensus"  
350..568  
/note="L2 repeat: matches 1..223 of consensus"  
570..1016  
/note="L1M2 repeat: matches 5168..5617 of consensus"  
1119..1184  
/note="L2 repeat: matches 2210..2276 of consensus"  
1189..1327  
/note="L1M1B repeat: matches 212..389 of consensus"  
1328..1393  
/note="33 copies 2 mer tt 72% conserved"  
1394..2645  
/note="L1M2 repeat: matches 4895..5146 of consensus"  
2679..2702  
/note="L1P repeat: matches 4898..4921 of consensus"  
2727..3230  
/note="L1M9 repeat: matches 5724..6270 of consensus"  
3290..3329  
/note="L2 repeat: matches 2575..2614 of consensus"  
4634..4810  
/note="MIR repeat: matches 25..212 of consensus"  
5377..5562  
/note="MIR repeat: matches 60..262 of consensus"  
7088..7654  
/note="L2 repeat: matches 2088..2750 of consensus"  
7653..7704  
/note="L2 repeat: matches 2698..2750 of consensus"  
7895..8052  
/note="MERS5B repeat: matches 187..341 of consensus"  
9172..9925  
/note="L1 repeat: matches 4570..5321 of consensus"  
11654..12021  
/note="M1T1 repeat: matches 41..410 of consensus"  
12285..12562  
/note="L1M2 repeat: matches 1..289 of consensus"  
12563..12913  
/note="L1M3 repeat: matches 1..354 of consensus"  
12914..14507  
/note="L1M4 repeat: matches 1..1580 of consensus"  
14508..14850  
/note="L1M5 repeat: matches 1..354 of consensus"  
14851..15021  
/note="L1M6 repeat: matches 289..434 of consensus"  
16263..16372  
/note="L2 repeat: matches 2627..2746 of consensus"  
17204..17573  
/note="L1M7 repeat: matches 1..371 of consensus"  
18729..18975  
/note="L1M8 repeat: matches 2172..2418 of consensus"  
19521..19829

repeat\_region  
/note="L1M1 repeat: matches 6..311 of consensus"  
20023..20996  
/note="L1M2 repeat: matches 274..1293 of consensus"  
21916..21981  
/note="L2 repeat: matches 2683..2749 of consensus"  
22157..22563  
/note="M1T1 repeat: matches 1..394 of consensus"  
24729..25549  
/note="L1M3 repeat: matches 6676..7531 of consensus"  
26030..26119  
/note="L2 repeat: matches 2627..2710 of consensus"  
26692..26981  
/note="L1M4 repeat: matches 1..294 of consensus"  
29249..33342  
/note="L1M5 repeat: matches 4..6143 of consensus"  
37018..37075  
/note="29 copies 2 mer ta 79% conserved"  
37670..37876  
/note="MIR repeat: matches 56..250 of consensus"  
39156..39294  
/note="L1M6 repeat: matches 3246..3378 of consensus"  
39833..39898  
/note="L1M7 repeat: matches 459..566 of consensus"  
40670..40894  
/note="L1M8 repeat: matches 2313..2543 of consensus"  
40948..41068  
/note="L1M9 repeat: matches 1..121 of consensus"  
43793..43955  
/note="L1M10 repeat: matches 5073..5239 of consensus"  
43992..45055  
/note="L1M11 repeat: matches 3920..5055 of consensus"  
45062..45407  
/note="L1M12 repeat: matches 2266..2621 of consensus"  
45433..45823  
/note="L1M13 repeat: matches 2636..3033 of consensus"  
45834..46449  
/note="L1M14 repeat: matches 5063..5678 of consensus"  
46611..46714  
/note="L1M15 repeat: matches 5774..6288 of consensus"  
46771..47268  
/note="L1M16 repeat: matches 3035..3259 of consensus"  
47298..47513  
/note="L1M17 repeat: matches 1..251 of consensus"  
47783..48349  
/note="L1M18 repeat: matches 396..950 of consensus"  
48356..48070  
/note="L1M19 repeat: matches 1..711 of consensus"  
49251..49608  
/note="L1M20 repeat: matches 5189..5550 of consensus"  
49667..49757  
/note="L1M21 repeat: matches 5557..5647 of consensus"  
51051..51304  
/note="L1M22 repeat: matches 1..564 of consensus"  
52008..52150  
/note="L1M23 repeat: matches 2687..2750 of consensus"  
52980..53147  
/note="L1M24 repeat: matches 22..188 of consensus"  
53870..53960  
/note="L1M25 repeat: matches 50..140 of consensus"  
54155..54580  
/note="L1M26 repeat: matches 445..881 of consensus"  
54647..54826  
/note="L1M27 repeat: matches 8..197 of consensus"  
55219..56488  
/note="L1M28 repeat: matches 4863..6139 of consensus"  
57097..57418  
/note="L1M29 repeat: matches 5975..6291 of consensus"  
57424..57638  
/note="L1M30 repeat: matches 7168..7388 of consensus"  
57634..58012  
/note="L1M31 repeat: matches 6605..7008 of consensus"



Fri May 25 10:28:24 2001

QY 172 tatttcttaaccaatccagattgcacagtatagggttaagtaactcacagcaagaatctaact 231  
Db 172721 TATTGTGAAGATTCTCCAAATGCGACTTTAACTCCATAAAAAACACAAGAGAAATATTACAGATT 172662  
QY 232 gaaaaacccctcatatagaggtatatgaaacgycgaagaacgggagcaaaaaacagcaag 291  
Db 172601 TTAATAATACACGTTATGGAACAATCATATCAAGCATCAAAATAATGAGCCACAAACAAAGGACAG 172602  
QY 292 cccaatga 299  
Db 172601 CTGGATCA 172594

Search completed: May 25, 2001, 04:08:32  
Job time: 6453 sec

